Ekram Hossain
Kin K. Leung
*Editors*

# Wireless Mesh Networks

## Architectures and Protocols

Springer

# Wireless Mesh Networks

Architectures and Protocols

Ekram Hossain · Kin Leung
Editors

# Wireless Mesh Networks
## Architectures and Protocols

Springer

*Edited by:*

Ekram Hossain
Department of Electrical &
Computer Engineering
University of Manitoba
75A Chancellor's Circle
Winnipeg MB R3T 5V6
CANADA

Kin Leung
Imperial College
Department of Electrical &
Electronic Engineering
Exhibition Rd.
London SW7 2BT
UNITED KINGDOM

Library of Congress Control Number: 2007933940

ISBN: 978-0-387-68838-1 e-ISBN: 978-0-387-68839-8

Printed on acid-free paper.



9 8 7 6 5 4 3 2 1

springer.com

To our families

# Preface

**A Brief Journey through "Wireless Mesh Networks: Architectures and Protocols"**

*Ekram Hossain*, University of Manitoba, Winnipeg, Canada
*Kin Leung*, Imperial College, London, United Kingdom

## Introduction

Wireless mesh networking has emerged as a promising design paradigm for next generation wireless networks. Wireless mesh networks (WMNs) consist of mesh clients and mesh routers, where the mesh routers form a wireless infrastructure/backbone and interwork with the wired networks to provide multihop wireless Internet connectivity to the mesh clients. Wireless mesh networking has emerged as one of the most promising concept for self-organizing and auto-configurable wireless networking to provide adaptive and flexible wireless Internet connectivity to mobile users. This concept can be used for different wireless access technologies such as IEEE 802.11, 802.15, 802.16-based wireless local area network (WLAN), wireless personal area network (WPAN), and wireless metropolitan area network (WMAN) technologies, respectively. Potential application scenarios for wireless mesh networks include backhaul support for cellular networks, home networks, enterprise networks, community networks, and intelligent transport system networks. Development of wireless mesh networking technology has to deal with challenging architecture and protocol design issues, and there is an increasing interest on this technology among the researchers in both academia and industry. There are many on-going research projects in different universities and industrial research labs. Also, many startup companies are building mesh networking platforms based on off-the-shelf wireless access technologies and developing demanding applications and services. This book intends to provide a unified view of the state-of-the-art achievements in the area of protocols and architectures for wireless mesh networking technology.

The contributed articles in this book from the leading experts in this field cover different aspects of analysis, design, deployment, and optimization of protocols and architectures for WMNs. In particular, the topics include challenges and issues in designing architectures and protocols for WMNs, medium access control and routing protocols for WMNs, resource allocation and scheduling in WMNs, cost optimization in WMN nodes using energy harvesting technologies, cross-layer design for WMNs, and security in WMNs.

## Issues in Architecture and Protocol Design for Wireless Mesh Networks

*Chapter 1*, authored by *V. C. Gungor, E. Natalizio, P. Pace, and S. Avallone*, provides a comprehensive introduction to the recent developments in the protocols and architectures of wireless mesh networks (WMNs) and also discusses the opportunities and challenges of wireless mesh networks. The major issues related to wireless mesh network architecture and management include network planning (e.g., placement of mesh routers, number and type of network interfaces in each router), network integration (i.e., integration of WPAN, WLAN, and WMAN technologies), network scalability (i.e., ability to deal with large network topology), and flexible and scalable network management. The protocols for wireless mesh networks should be able to exploit the advanced wireless technologies (e.g., cognitive/reconfigurable radio, multiple-input multiple-output (MIMO) radio), provide quality of service (QoS) to different types of applications, provide efficient network self-reconfiguration, topology control, power management, provide mobility support, and provide mechanisms for efficient encryption, authentication, and intrusion detection.

The authors have described the major research issues at the different layers in the protocol stack of a wireless mesh network. At the application layer, new protocols need to be designed for distributed information sharing and to address the pricing and incentive issues. Again, the application layer protocols need to work in cohesion with the lower layer protocols to meet the application requirements in an efficient manner.

Efficient transport protocols would be required for non-real-time and real-time applications in wireless mesh networks. Due to the dynamic characteristics of multihop communication environment in a wireless mesh network as well as the integration of different types of networking technologies, the traditional transport protocols (e.g., TCP-based protocols) may experience significant performance degradation. In particular, under-utilization of network resources may result due to the increased round-trip time (RTT), large variance in RTT estimate, and increased link error rate in the network as well as the end-to-end congestion detection and control mechanisms used in these protocols. Again, since the traditional TCP-friendly rate control protocols for multimedia delivery handle all non-congestion-related packet losses in the same way, they would suffer performance inefficiency. Design of dynamic adaptive transport protocol for high performance real-time data transport and real-time

multimedia communications in wireless mesh networks is a grand research challenge.

For wireless mesh networks, simple (i.e., low overhead), scalable, distributed, load-balancing and link quality-aware routing protocols would be required for efficient multihop communications. Designing efficient routing protocols for multi-channel and multi-radio mesh networks is a major research challenge. An integrated design of routing, medium access control, and channel allocation (or scheduling) may lead to an efficient solution.

Multi-channel and multi-radio-aware MAC protocols are promising for wireless mesh networks. Channel allocation among multiple radios should be performed in a way so that the network connectivity is preserved and the co-channel interference remains below the acceptable limit while at the same time the maximum frequency reuse is achieved. Also, multi-rate transmission and adaptivity to dynamic network configuration are desirable.

High-speed physical layer techniques such as MIMO, beamforming and smart antennas, reconfigurable/cognitive radio will enable to increase the capacity and reliability of wireless mesh networks. These advanced physical layer techniques can be fully utilized by making the higher-layer protocols aware of the physical layer and using the low-cost software radio platform.

Specifications for wireless mesh networks are being standardized by the IEEE 802.11, IEEE 802.15, and IEEE 802.16 standard groups. 802.11s task group was set up by IEEE for installation, configuration, and operation of IEEE 802.11-based wireless mesh networks. IEEE 802.15.5 task group is working towards developing an architectural framework for mesh networking among IEEE 802.15-based WPAN devices. IEEE 802.16a standard for broadband wireless access in metropolitan area networks support mesh mode of operation for fixed broadband applications in which the subscriber stations can directly communicate with each other through multihop communications. The Mobile Multihop Relay (MMR) study group under the IEEE 802.16 working group is developing specifications for supporting mobile stations by using multihop relaying techniques through relay stations.

Recent field trials and experiments on wireless mesh networks (built from off-the-shelf wireless technologies) in several academic research testbeds and commercial installations have shown that the performance is not quite satisfactory. This reflects the need for development of novel architectures and protocol suites to address the issues such as QoS, scalability, heterogeneity, self-reconfuguration, and security for wireless mesh networks.

*Chapter 2*, authored by *J.-H. Huang, L.-C. Wang, and C.-J. Chang* first describes the major wireless mesh network architectures, namely, the backbone wireless mesh network, backbone with end-user wireless mesh network, and relay-based wireless mesh network architectures. In a wireless multihop backbone network, each of the base stations (or access points (APs)) operates as a relay to forward traffic from other base stations to the Internet gateway. In a backbone with end-user wireless mesh network, both the base stations and end users act as relays to forward traffic from neighboring nodes, and thereby, it improves the coverage of base stations and enhances network connectivity.

The authors address the scalability issue in wireless mesh networks from the network deployment perspective. The authors propose two scalable wireless mesh network deployment strategies, namely, cluster-based wireless mesh and ring-based wireless mesh for dense urban coverage and wide-area coverage scenarios, respectively. In a cluster-based wireless mesh, several adjacent access points, which are connected "wirelessly", form a cluster and only one of the access points connects to the Internet. The ring-based wireless mesh is based on a mesh cell architecture where the cell is divided into several rings allocated with different channels. The central gateway (which is connected to the Internet) and the stationary mesh nodes in the cell form a multihop wireless mesh network. The authors investigate the tradeoff between capacity and coverage for these two scalable wireless mesh architectures. With multiple available channels, the scalability can be improved through proper frequency planning and proper design of the deployment parameters in these networks. Note that, while a larger cell size is preferred from the coverage viewpoint, a smaller cell size would be preferable to achieve a higher data rate.

The authors apply a mixed-integer nonlinear programming (MINLP)-based optimization approach to determine the optimal deployment parameters (i.e., separation distance for access points in a cluster-based wireless mesh network) under given coverage and rate constraints where the objective is to maximize the ratio of total offered traffic load to the cost for a cluster of access points. Two AP placement strategies, namely, the increasing-spacing and the uniform-spacing strategies are considered. In case of increasing-spacing placement strategy, access points are deployed with increasing separation distance from the central access point. In case of uniform-spacing placement strategy, all the cells in a cluster have the same radius. Numerical results show that the increasing-spacing strategy outperforms the uniform-spacing strategy and there exists an optimal value of the number of access points which maximizes the objective function.

For the ring-based wireless mesh network, an MINLP formulation is used to determine the optimal number of rings in a cell and the optimal width of each ring for which the desired tradeoff between throughput and coverage can be achieved. Numerical results assuming IEEE 802.11a-based wireless access show that the ring-based wireless mesh improves both the coverage and the cell throughput significantly compared to the single-hop network.

## Information Theoretic Characterization of End-to-End Performance in Cellular Mesh Networks

*Chapter 3*, authored by *Ö. Oyman and S. Sandhu*, provides results on information-theoretic characterization of end-to-end performance in terms of physical channel and system parameters in an orthogonal frequency division multiplexing (OFDM)-based multihop cellular mesh network. Specifically, the capacity is defined as the end-to-end (instantaneous) conditional mutual information which is a function of the random fading channel parameters and the transmit signal-to-noise ratio. This conditional mutual information can be computed for each hop considering practical

link adaptation mechanisms based upon which an end-to-end link quality parameter can be obtained.

Through simulation, the authors demonstrate that, for users at the edge of a cell, multihop relaying can provide capacity and coverage gains over direct transmission. Also, multihop relaying improves the end-to-end capacity compared to single-hop communication, specially at the low outage probability regime. The optimal number of hops, which maximizes the end-to-end mutual information is observed to be sensitive to the channel parameters.

Based on a Markov chain model, the authors also analyze the end-to-end throughput and latency over a multihop network which supports automatic repeat request (ARQ)-based error control at each hop along a routing path. Based on this analysis, the routing metric at each hop can be obtained, and subsequently, the throughput-maximizing (or latency-minimizing) routing path can be determined.

To this end, the authors present a centralized resource allocation framework for user scheduling, subcarrier allocation, and multihop route selection in orthogonal frequency division multiple access (OFDMA)-based relay-assisted cellular mesh networks. In this framework, the base station decides on the allocation of time and frequency resources across users and it also coordinates the actions of the relay terminals. To reduce system design complexity, multihop route selection and subcarrier allocation are performed separately. The link quality metrics are used to choose the multihop routing paths for each user such that the end-to-end capacity is maximized. The end-to-end route metrics for all users over all subcarriers are then used for scheduling the subcarriers. The authors also demonstrate how the information theoretic analysis of end-to-end capacity can be used to determine the optimal policies for network entry and handoff.

## Medium Access Control and Routing Protocols for Wireless Mesh Networks

*Chapter 4*, authored by *J. C. Hou, K.-J. Park, T.-S. Kim, and L.-C. Kung*, provides a comprehensive survey on the state of the art in design and implementation of medium access control (MAC) and routing protocols for wireless mesh networks. The objective of a MAC protocol in such a network is to maximize network capacity (e.g., through improved spatial reuse) while providing required quality of service (QoS) performances to the users. The major issues related to MAC design in a wireless mesh network are - controlling the sharing range of the wireless medium and increasing spatial reuse, exploiting availability of multiple channels, and exercising rate control. The spatial reuse can be improved by either reducing the transmit power (while maintaining network connectivity) or increasing the carrier sense threshold (while mitigating MAC-level interference). Capacity improvements can be achieved by using multiple radio interfaces in each mesh node where orthogonal channels are assigned to the radios. Distributed dynamic assignment of channels among the radios as well as joint optimization of routing and channel assignment are challenging

research problems. Network throughput can be maximized through dynamic adaptation of data rate according to the channel condition, that is, by selecting the highest possible data rate for a given signal-to-interference-plus-noise ratio (SINR) that allows correct decoding of packets at the receiver.

The authors summarize the related works on transmit power control, carrier sense adaptation, and exploiting spatial-temporal diversity which are intended to improve the spatial reuse/capacity of the network. In the literature, the transmit power control problem in wireless ad hoc/sensor/mesh networks has been studied by using graph-theoretic approaches in the context of topology maintenance. The major objective here is to mitigate MAC interference while preserving network connectivity. The graph-model-based topology control algorithms aim at keeping the node degree in the communication graph low with the assumption that low node degree implies low interference. However, in a graph model, since node degree may not adequately capture the physical interference, graph-model-based topology control may result in low network capacity and volatile network connectivity. There have been other approaches for transmit power control which aim at maximizing network capacity.

A number of studies in the literature focused on adaptation of carrier sense threshold to improve the level of spatial reuse. The selection of the optimal carrier sense threshold depends on the factors such as the SINR threshold, level of channel contention (i.e., traffic load), transmit power, network topology, hidden/exposed nodes, type of flows (i.e., single hop or multihop), bidirectional handshakes, packet size, and MAC overhead. The relationship between the transmit power and the carrier sense threshold impact network capacity. For example, with low transmit power and high carrier sense threshold, a large number of concurrent transmissions can be supported, with each transmission sustaining a low data rate. Several works in the literature addressed the problem of joint control of transmit power and carrier sense threshold. Again, transmit power can be jointly optimized with rate control to maximize network capacity. For a rate-adaptive MAC protocol, data rate is generally increased/decreased on consecutive transmission success/packet loss. The rate control problem at the MAC layer has been studied quite extensively in the literature.

In a wireless mesh network, the spatial diversity that exists among the multihop paths, can be exploited to improve network capacity. Again, capacity can be improved through multi-channel and multi-radio design for wireless mesh networks. Specifically, in the MAC layer, multiple channels can be exploited to achieve higher throughput as well as to mitigate the fairness problem in a multihop environment. Multiple radios in a node enable it to communicate with other nodes in a full-duplex manner with minimal interference.

The major objective of a routing protocol for wireless mesh networks is to determine high-throughput routes (i.e., interference-mitigated routes) between nodes so that the maximal end-to-end throughput can be achieved. Instead of using the conventional hop-count-based route metric, link quality-based route metrics have been proposed for routing in wireless mesh networks. In the literature, routing protocols have been proposed for single-radio single-channel, single-radio multi-channel, and multi-radio multi-channel wireless mesh networks. In a multi-channel and multi-radio mesh network, by properly assigning the different channels to the different ra-

dios, intra- and inter-flow interference can be avoided and interference-free/mitigated routes can be constructed.

To this end, the authors introduce a modular programming environment to enable cross-layer design and optimization in wireless mesh networks. In this environment, physical layer (PHY)/MAC parameters and events can be exported to higher-layer protocol modules. Controlled transparency, flexibility, and easy integration and portability are some of the features of this programming environment.

## Channel Assignment Strategies in Multi-channel and Multi-radio Wireless Mesh Networks

*Chapter 5*, authored by *M. Conti, S. K. Das, L. Lenzini, and H. Skalli*, deals with the problem of assigning channels to radio interfaces in a multi-channel and multi-radio wireless mesh backbone network. The key challenges associated with the channel assignment problem are outlined and a survey on the existing channel assignment schemes is provided.

The objective of a channel assignment strategy is to ensure efficient utilization of the available channels (e.g., by minimizing interference) while maximizing connectivity in the network. However, since these two requirements are conflicting with each other, the goal is to achieve a balance between these two. The major constraints which need to be satisfied by a channel assignment scheme include: fixed number of channels in the network, limited number of radios in a mesh node/router, common channel between two communicating nodes, and limited channel capacity. Also, a channel assignment scheme should take the amount of traffic load supported by each mesh node into consideration.

Optimal channel assignment in an arbitrary wireless mesh backbone is an NP-hard problem (similar to the graph coloring problem). The existing channel assignment schemes in the literature are, therefore, mostly heuristic based. These schemes can be classified into three categories: fixed, dynamic, and hybrid channel assignment schemes. Fixed assignment schemes assign channels to the radios either permanently or for a long time interval. With dynamic channel assignment, the radios can frequently switch from one channel to another. Hybrid channel assignment strategies apply a fixed assignment for some radios and a dynamic assignment for other radios.

Fixed channel assignment schemes can be further classified into two categories: common channel assignment (CCA) schemes and varying channel assignment (VCA) schemes. In CCA, all the radios in all of the mesh nodes are assigned the same set of channels. In VCA, radios of different nodes are assigned different sets of channels. The authors have described a number of such VCA schemes.

With dynamic channel assignment, when two mesh nodes need to communicate with each other, they need to switch to the same channel. The key challenge in this case is how to coordinate the switching decisions. The authors have described a number of dynamic channel assignment schemes.

Hybrid assignment strategies are attractive since they allow for simple coordination algorithms (as for the fixed assignment schemes) and also provides the flexibility

of dynamic channel assignment. The authors have described two such hybrid channel assignment schemes.

The key issues considered in the design of the existing channel assignment schemes are network connectivity, constraint on topology, interference minimization, effects of link revisits, traffic awareness, switching overhead (for dynamic and hybrid schemes), and control philosophy (i.e., centralized or distributed). Considering these factors, the authors provide a qualitative comparison among the different schemes.

## Resource Allocation for Wireless Mesh Networks

### Resource Allocation and Transmission Rate Control

*Chapter 6*, authored by *Y. Xue, Y. Cui, and K. Nahrstedt*, presents a generalized theoretical framework for resource allocation and transmission rate control in wireless mesh networks. The objective of this framework is to achieve optimal resource utilization and rate fairness among flows on an end-to-end basis. Based on this theoretical framework, the authors also present a price-based distributed algorithm for resource allocation which converges to the globally optimal solution.

The resource allocation problem is first formulated as an optimization problem for an abstract network model consisting of a set of resource elements (e.g., wireless links) which are shared by a set of flows. The objective is to maximize the aggregated utility (i.e., satisfaction) for all flows under constraints on capacities of the resource elements. Different fairness models such as weighted proportional fairness and max-min fairness can be implemented through the appropriate choice of the utility function. The solution of the optimization achieves both optimal resource utilization (i.e., Pareto optimal rate allocation) and fair allocation of transmission rate among end-to-end flows. Based on the Lagrangian form of the optimization formulation, a price-based decentralized solution can be obtained which depends on local decision of each resource element and exchange of control signals among them.

The authors show that for a multihop wireless mesh backbone network, a resource element is a facet of the polytope defined by the independent sets of the conflict graph of this network. It can be approximated by a maximal clique of the contention graph which basically represents a maximal distinct contention region in the network. The resource constraints in the network can then be represented by the achievable channel capacities in all of the maximal cliques in the contention graph. Subsequently, the end-to-end rate allocations can be obtained for the flows. For distributed implementation, a flow adapts its rate as a function of price it pays to all resource elements, where the price for a resource element is a non-negative, continuous, and increasing function of the total traffic served by that resource element. The authors show that the rate adaptation algorithm is stable and at the equilibrium each flow maximizes its utility.

### Resource Allocation in Solar/Wind-Powered Mesh Nodes

*Chapter 7*, authored by *A. A. Sayegh, T. D. Todd, and M. N. Smadi*, presents some experimental results on resource allocation in hybrid solar/wind powered WLAN mesh nodes. Resource allocation in such a node involves assigning solar panel or wind turbine size, and battery capacity, and this resource allocation depends on the geographic location of the node. A sustainable energy WLAN mesh node includes a wind turbine and/or solar panel which are connected to a battery through a charge controller. The charge controller disconnects the battery from the power source to protect it from under- and over- charging. Specifically, when the residual battery energy falls below the maximum allowed level of discharge, the charge controller disconnects the node load and the node then experiences a radio outage. In a hybrid configuration, both solar panel and wind turbine are used.

The authors investigate the short-term statistics of the energy available from solar panel and wind turbines at two different locations, namely, Toronto, Ontario and Phoenix, Arizona. In the city of Toronto, a time distribution example of solar power and wind power shows positive correlation between them which suggests that a hybrid solar/wind powered node may not be cost effective. In the city of Phoenix, comparison of solar power and wind power shows that solar power dominates the wind power, and therefore, wind power alone or a hybrid wind/solar solution may not be feasible. However, the short-term statistics may not be sufficient to assess the optimal dimensioning of the power source in the mesh node. The long-term statistics would be required instead. Examples of long-term statistics show that performance metrics such as radio outage probability for the wind source and the solar source depends on the seasonal correlation between solar power and wind power in a geographic location. The desired level of sustainability of a given hybrid system for the different geographical locations can be obtained by properly choosing the wind turbine and battery sizes.

To minimize the total cost of a hybrid node (i.e., cost of battery, solar panel, and wind turbine) under given constraints on outage probability, battery size, solar panel and wind turbine size, the authors use an optimization formulation. This optimization model is solved numerically. To this end, the authors show that power saving at mesh access points can greatly reduce the cost which is almost proportional to the power consumption in the node.

## Scheduling, Routing, and Cross-Layer Design

### Link Scheduling and Routing in Wireless Mesh Networks

*Chapter 8*, authored by *L. Badia, A. Erta, L. Lenzini, and M. Zorzi*, presents a comprehensive survey on the state-of-the art of routing and link scheduling in wireless mesh networks. As has been mentioned before, for a wireless mesh network, the objective of a routing algorithm is to discover efficient paths to obtain high system throughput. Link scheduling at the medium access control layer is used to activate

the communication links with an objective to ensuring the desired level of network connectivity under interference constraints. The interference models, which are particularly important when designing link scheduling (or activation) and routing algorithms, can be of three types - *physical*, *protocol*, and *measurement-based* interference models. With a *physical* interference model, the feasibility of simultaneous link activations is determined by the SINR at the receivers. Note that, the packet error rate at a receiver is a monotonically decreasing function of SINR. With a *protocol* interference model, simultaneous transmissions result in incorrect decoding of a received packet. The measurement-based interference model takes an *a priori* approach to interference characterization.

The existing works on link scheduling and routing in wireless ad hoc and/or sensor networks are often not suitable in the context of wireless mesh networks due to the dissimilar design/optimization goals and/or oversimplified interference models. Designing a framework for joint scheduling and routing which considers the network requirements, resource constraints (e.g., number of radios, channels), radio transceiver constraints, and realistic interference models is an interesting research challenge.

The authors propose a graph-based approach to design a framework for joint link scheduling and routing through link activation. In this framework, the radio transceiver constraints (e.g., half-duplexity) and link directionality are taken into account. The interference is characterized by a physical interference model which is more accurate than that under protocol interference models from the viewpoint of theoretical analysis of wireless mesh networks. The authors assume a centralized space time division multiple access (STDMA) scheme to obtain an efficient transmission scheme through link activation. The mesh access point nodes in the mesh backbone network finds the link activation patterns in a centralized manner and communicates it with the other nodes. The authors obtain the performance bounds for the minimal time scheduling problem/shortest-time link activation pattern (i.e., obtaining the link activation pattern which delivers a given amount of traffic from non-gateway mesh nodes to the gateway mesh nodes in the shortest possible time). The authors also carry out some numerical investigations on the performance of the proposed framework for different interference models.

**Quality-Aware Routing Metrics in Wireless Mesh Networks**

*Chapter 9*, authored by *C. E. Koksal*, presents a comparative study among seven different link cost metrics for routing in wireless mesh networks. The cost metric for a link refers to the cost of forwarding a packet along that link. The considered link cost metrics are: hop count, per-hop round trip time (RTT), per hop packet pair delay (PktPair), quantized loss rate, expected transmission count (ETX), modified ETX (mETX), and effective number of transmissions (ENT).

The traditional hop count-based routing (i.e., minimum hop routing), although simple and requires minimal amount of measurement, does not perform satisfactorily in presence of link variability. Per-hop round trip time is a delay-based link cost metric, which is calculated by a mesh node as the exponentially weighted moving

average of the RTT samples for each of its neighbors. This metric takes into account the factors such as queueing delay, channel quality, and channel contention. However, since RTT varies with varying load, using this routing metric may lead to route instability (due to the self interference effect). With this routing metric, the optimal path assignments may change more frequently compared to the hop count, which may result in reduced network throughput. Also, this metric responds to channel variability at time scales longer than tens of packets.

The PktPair metric is obtained as the difference between the times of reception of two successive packets. Therefore, it does not take into account the queueing and processing delay at a node. Although it suppresses the route instability effect to some extent, the overhead associated with it is higher than that due to per hop RTT. The quantized loss rate is based on the end-to-end packet loss probability. This metric does not take the link bandwidth into account, and therefore, low bandwidth paths could be chosen for routing.

ETX for a wireless link refers to the estimated expected number of transmissions required to transfer a packet successfully over that link. This metric depends only on the link level packet errors due to channel impairments, and therefore, the effects of self interference is reduced. ETX can improve the throughput performance significantly compared to the hop count metric, however, it may perform poorly under highly variable and bursty error situations. The mETX metric overcomes the limitations of ETX in the presence of channel variability. This metric is a function of the mean and the variance of the bit error probability summed over a packet duration. It offers a higher throughput performance compared to the ETS metric. However, the main drawback of this metric is the complexity of estimation of the mean and variance of bit error probability. Also, estimation error may impact its performance significantly.

The ENT metric is structurally similar to the mETX metric and it uses the exactly same parameters and the channel estimation procedure as mETX. It is used to find routes which satisfy certain desired end-to-end performance (e.g., packet loss rate at the transport level) requirements. The metric mETX can be considered as a special case of the ENT metric.

The authors also present a unified geometric framework to compare the different routing metrics. This framework combines the mean and standard deviation of the bit error rate process. In this framework, it is possible to define a feasible region using which links can be selected to achieve the desired routing performance.

### Cross-Layer Solutions for Traffic Forwarding in Wireless Mesh Networks

*Chapter 10*, authored by *V . Baiamonte, C. Casetti, C. F. Chiasserini, and M. Fiore*, deals with the problem of joint design of MAC and routing schemes for multihop communication in IEEE 802.11-based wireless mesh networks. Specifically, the authors consider the problem of designing efficient relaying schemes based on the cross-layer design principles which take into account the quality of the wireless links in an 802.11-based multi-rate WLAN.

For traffic flow from a mesh gateway to wireless mesh nodes, the authors present two schemes for packet forwarding, namely, the split queues (SQ) approach and the access category (AC) approach. With the former approach, two queues are maintained at each node for relay traffic and local traffic. With the latter approach, several queues are implemented at the MAC layer, each of which is associated with a priority level (e.g., implementable through the access categories defined in IEEE 802.11e EDCA). Prioritizing relay traffic over local traffic provides an incentive to the nodes to act as relays. Simulation results for a network topology with single and multiple relays serving TCP and UDP flows show that the AC approach can provide significant gain in throughput while the SQ approach can provide very high fairness in throughput.

The authors present a fair relay selection algorithm (FRSA) which is an extension of the optimized link state routing (OLSR) protocol designed for wireless ad hoc networks. OLSR is a table-driven and a proactive protocol which exchanges topology information periodically with other nodes in the network. The route from a given node to any destination node in the network is formed by relay nodes. A relay node announces to the network that it has reachability to the nodes which have selected it as the relay node. The proposed FRSA is a relay quality-aware routing extension of OLSR. In FRSA, each node performs a relay quality-aware routing to its two-hop neighborhood. Simulation results show that a significant throughput gain with fair channel access can be achieved with FRSA when compared to OLSR.

## Multiple Antenna Techniques for Wireless Mesh Networks

*Chapter 11*, authored by *A. Gkelias and K. K. Leung*, discusses the research challenges associated with the deployment of multiple antenna technologies in wireless mesh networks. In particular, the authors focus on the design of medium access control and routing algorithms in wireless mesh networks employing smart antenna technology. Multiple antenna technology includes fixed beam antenna techniques, adaptive antenna techniques, and multiple-input multiple-output (MIMO) coding techniques which can be highly beneficial to improving overall performance of wireless mesh networks. However, employment of multiple antenna (or smart antenna) techniques in a wireless mesh networking environment gives rise to unique problems such as deafness, hidden and exposed terminals, and multi-stream interference. Novel medium access control and routing protocols need to be designed to address the above problems.

The authors first describe the wireless mesh network and channel characteristics considering different propagation scenarios, interference characteristics in different scenarios, and other constraints such as the limitations in total effective radiation power. Then an overview of the different smart antenna techniques is provided. Two basic types of smart antennas, namely, directional antennas (fixed beams) and adaptive antenna arrays, are considered. Directional antenna techniques, which include switched-beam antennas, steered-beam antennas (or dynamically phased array antennas), can provide high SINR gain in presence of strong line-of-sight component,

however, their performances degrade in multi-path environments. Adaptive antenna techniques, which include adaptive antenna arrays and MIMO techniques, can provide high gain in the direction of desired signals and nulls in the direction of undesired signals (i.e., interference). In particular, the MIMO techniques can exploit the multi-path fading effects to enhance the transmission rate (i.e., multiplexing gain) or enhance the transmission reliability (i.e., diversity gain) without additional bandwidth requirements.

One of the major issues related to the use of multiple antenna (or smart antenna) techniques in wireless mesh networks is to mitigate the deafness problem. This problem arises due to the use of directional antennas when a transmitter fails to communicate with its intended receiver. However, deafness can be also exploited in some cases to mitigate interference. Directional transmission may also augment the classical hidden/exposed terminal problem in wireless networks. Again, in presence of directional antennas, unsuccessful transmissions due to packet collision and deafness need to be treated differently at the higher layers. In a MIMO-based wireless mesh network, the medium access control protocol should use the optimal number of simultaneous transmissions, allocate appropriate number of streams per transmitter-receiver pair, and perform power allocation accordingly. Also, the tradeoff between multiplexing and diversity gain should be taken into account. The routing protocols in a MIMO-based wireless mesh network should consider the MIMO parameters for route discovery and maintenance. If the higher layer protocols are not carefully designed, the multiple antenna techniques can have negative impact on the overall network performance.

The authors then discuss several distributed medium access control protocols for multiple antenna-based multihop wireless networks. The interactions between medium access and routing protocols in presence of smart antennas have been evaluated in some works in the literature. These works primarily focused in improving network connectivity. Design and implementation of efficient quality of service (QoS)-aware routing protocols which exploit the multiple antenna techniques is a grand research challenge.

## Security in Wireless Mesh Networks

*Chapter 12*, authored by *W. Zhang, Z. Wang, S. K. Das, and M. Hassan*, addresses the security issues in wireless mesh networks. The main challenges for securing wireless mesh networks arise due to the requirements of authentication, secure routing, secure location information (of mesh routers), and to defend against virus attacks.

Authentication is required to distinguish malicious information from legitimate information. An authentication mechanism is generally implemented with the help of public key infrastructure (PKI) and certification authority (CA). With the PKI mechanism, each user has a pair of cryptographic keys: public key and private key. A message encrypted with the public key (which is known to all the users) can only be decrypted by using the corresponding private key, and vice versa. The CA involved in the authentication procedure signs the binding of an entity's identity and its public

key with its private key. It is assumed that the signed certificates by the CA are globally trusted in the network. Due to the absence of any pre-established trusted network infrastructure in wireless mesh networks, distributed CA schemes are desirable. The authors describe a number of such CA schemes.

The routing protocols for a wireless mesh network are vulnerable to both external and internal attacks. External attackers can inject fabricated routing information into the network or maliciously alter the contents of routing messages. An internal attack is launched from within a node when an attacker gains full control of the node. To prevent external attackers from sending fabricated routing information, cryptography-based authentication methods incorporated in the routing protocols can be used. The authors describe several of such schemes. Also, several possible approaches to detect and counter measure the internal attacks to routing protocols are discussed.

Securing the location information of wireless mesh routers is crucial for certain type of routing schemes (e.g., geographic routing schemes). Two methods for securing location information are generally used - correctly computing the location information and verifying the location claims. The authors review several works based on these two methods.

Computer viruses also pose threats to security in wireless mesh networks. There have been research efforts towards modeling the virus propagation problem in wireless networks. Epidemic theory used in Biology is one popular technique used to investigate the virus spreading problem. Two schemes which use Epidemic theory to model the propagation of viruses and compromised nodes, respectively, are discussed.

The authors also outline a number of security-related research issues in wireless mesh networks. These include securing the medium access control protocols, defending against denial of service (DoS) attacks at the different layers in the protocol stack, designing cross-layer framework for self-adapted security mechanisms, customizing the security schemes based on the type of network (in a heterogeneous wireless mesh environment), and trust establishment and management. All of these issues represent fertile areas of future research in wireless mesh networks.

## Conclusion

We have provided a summary of the contributed articles in this book. We hope this summary would be helpful to follow the rest of the book easily. We believe that the readers will find the rich set of references in each of the articles very valuable. We would like to express our sincere appreciation to all of the authors for their excellent contibutions and their patience during the publication process of the book. We hope this book will be useful to both researchers and practitioners in this emerging area.

# Contents

# List of Contributors

**V. C. Gungor, E. Natalizio, P. Pace, and S. Avallone**
Georgia Institute of Technology, USA; University of Calabria, Italy; University of Calabria, Italy; University of Napoli, Italy

gungor@ece.gatech.edu
enatalizio@deis.unical.it
ppace@deis.unical.it
stavallo@unina.it

**J.-H. Huang, L.-C. Wang, and C.-J. Chang**
Department of Communication Engineering, National Chiao-Tung University, Taiwan, R.O.C.

hjh@mail.nctu.edu.tw
lichun@cc.nctu.edu.tw
cjchang@cc.nctu.edu.tw

**Ö. Oyman and S. Sandhu**
Intel Corporation, USA

ozgur.oyman@intel.com
sumeet.sandhu@intel.com

**J. C. Hou, K.-J. Park, T.-S. Kim, and L.-C. Kung**
Department of Computer Science, University of Illinois, Urbana, IL 61801, USA

jhou@cs.uiuc.edu
kjp@cs.uiuc.edu
tskim@cs.uiuc.edu
kung@cs.uiuc.edu

**M. Conti, S. K. Das, L. Lenzini, and H. Skalli**
Istituto di Informatica e Telematica (IIT),
Italian National Research Council (CNR), Italy;
Department of Computer Science and Engineering
The University of Texas at Arlington, USA;
Department of Information Engineering, University of Pisa, Italy;
Department of Computer Science and Engineering, IMT Lucca Institute for High Studies, Italy

Marco.Conti@iit.cnr.it
das@cse.uta.edu
l.lenzini@iet.unipi.it
habiba.skalli@imtlucca.it

**Y. Xue, Y. Cui, and K. Nahrstedt**

Vanderbilt University, USA; University of Illinois at Urbana-Champaign, USA

yuan.xue@vanderbilt.edu
yi.cui@vanderbilt.edu
klara@cs.uiuc.edu

**A. A. Sayegh, T. D. Todd, and M. N. Smadi**
McMaster University, Canada

todd@mcmaster.ca

**C. E. Koksal**
The Ohio State University, USA

koksal.2@osu.edu

**L. Badia, A. Erta, L. Lenzini, and M. Zorzi**
IMT Lucca Institute for Advanced Studies, Italy;
Dept. of Information Engineering, University of Pisa, Italy;
Dept. of Information Engineering, University of Padova, Italy

l.badia@imtlucca.it
a.erta@imtlucca.it
l.lenzini@iet.unipi.it
zorzi@dei.unipd.it

**V. Baiamonte, C. Casetti, C. F. Chiasserini, and M. Fiore**
Dipartimento di Elettronica, Politecnico di Torino, Italy

valeria@tlc.polito.it
claudio@tlc.polito.it
carla@tlc.polito.it
marco@tlc.polito.it

**A. Gkelias and K. K. Leung**
Department of Electrical and Electronic Engineering, Imperial College, UK

a.gkelias@imperial.ac.uk
kin.leung@imperial.ac.uk

**W. Zhang, Z. Wang, S. K. Das, and M. Hassan**
Department of Computer Science and Engineering,
The University of Texas at Arlington, USA;
School of Computer Science and Engineering,
University of New South Wales, Australia

wzhang@cse.uta.edu
das@cse.uta.edu
zhewang@cse.unsw.edu.au
mahbub@cse.unsw.edu.au

# 1

# Challenges and Issues in Designing Architectures and Protocols for Wireless Mesh Networks

V. C. Gungor[1], E. Natalizio[2], P. Pace[3], and S. Avallone[4]

[1] Georgia Institute of Technology, USA
gungor@ece.gatech.edu
[2] University of Calabria, Italy
enatalizio@deis.unical.it
[3] University of Calabria, Italy
ppace@deis.unical.it
[4] University of Napoli, Italy
stavallo@unina.it

## 1.1 Introduction

Wireless Mesh Network (WMN) is a promising wireless technology for several emerging and commercially interesting applications, e.g., broadband home networking, community and neighborhood networks, coordinated network management, intelligent transportation systems. It is gaining significant attention as a possible way for Internet service providers (ISPs) and other end-users to establish robust and reliable wireless broadband service access at a reasonable cost. WMNs consist of mesh routers and mesh clients as shown in Fig. 1.1. In this architecture, while static mesh routers form the wireless backbone, mesh clients access the network through mesh routers as well as directly meshing with each other.

Different from traditional wireless networks, WMN is dynamically self-organized and self-configured. In other words, the nodes in the mesh network automatically establish and maintain network connectivity. This feature brings many advantages for the end-users, such as low up-front cost, easy network maintenance, robustness, and reliable service coverage. In addition, with the use of advanced radio technologies, e.g., multiple radio interfaces and smart antennas, network capacity in WMNs is increased significantly. Moreover, the gateway and bridge functionalities in mesh routers enable the integration of wireless mesh networks with various existing wireless networks, such as wireless sensor networks, wireless-fidelity (Wi-Fi), and WiMAX [3]. Consequently, through an integrated wireless mesh network, the end-users can take the advantage of multiple wireless networks. Some of the benefits and characteristics of wireless mesh networks are highlighted as follows:

- **Increased Reliability:** In WMNs, the wireless mesh routers provide redundant paths between the sender and the receiver of the wireless connection. This eliminates single point failures and potential bottleneck links, resulting in significantly

increased communications reliability [3]. Network robustness against potential problems, e.g., node failures, and path failures due to RF interferences or obstacles, can also be ensured by the existence of multiple possible alternative routes. Therefore, by utilizing WMN technology, the network can operate reliably over an extended period of time, even in the presence of a network element failure or network congestion.

- **Low Installation Costs:** Recently, the main effort to provide wireless connection to the end-users is through the deployment of 802.11 based Wi-Fi Access Points (APs). To assure almost full coverage in a metro scale area, it is required to deploy a large number of access points because of the limited transmission range of the APs. The drawback of this solution is highly expensive infrastructure costs, since an expensive cabled connection to the wired Internet backbone is necessary for each AP. On the other hand, constructing a wireless mesh network decreases the infrastructure costs, since the mesh network requires only a few points of connection to the wired network. Hence, WMNs can enable rapid implementation and possible modifications of the network at a reasonable cost, which is extremely important in today' s competitive market place.
- **Large Coverage Area:** Currently, the data rates of wireless local area networks (WLANs) have been increased, e.g., 54 Mbps for 802.11a and 802.11g, by utilizing spectrally efficient modulation schemes. Although the data rates of WLANs are increasing, for a specific transmission power, the coverage and connectivity of WLANs decreases as the end-user becomes further from the access point. On the other hand, multi-hop and multi-channel communications among mesh routers and long transmission range of WiMAX towers deployed in WMNs can enable long distance communication without any significant performance degradation.
- **Automatic Network Connectivity:** Wireless mesh networks are dynamically self-organized and self-configured. In other words, the mesh clients and routers automatically establish and maintain network connectivity, which enables seamless multi-hop interconnection service. For example, when new nodes are added into the network, these nodes utilize their meshing functionalities to automatically discover all possible routers and determine the optimal paths to the wired Internet [3]. Furthermore, the existing mesh routers reorganize the network considering the newly available routes and hence, the network can be easily expanded.

In this chapter, we present a survey of recent developments in the protocols and architectures for WMNs and discuss the opportunities and challenges of WMNs. The motivation of this chapter is to provide a better understanding of wireless mesh network technology that can ensure heterogeneous application requirements. Consequently, our aim is to present a structured framework for the end-users who plan to utilize WMNs for their applications and hence, to make the decision-making process more effective and direct.

The rest of the chapter is organized as follows. In Section 1.2 the network architecture of WMNs is presented, while in Section 1.3 the design challenges of WMNs are described. The recent advances and open research issues in protocol design for

WMNs are investigated in Section 1.4. Physical testbeds and standardization activities in WMNs are explored in Section 1.5 and 1.6, respectively. Finally, the conclusions are stated.

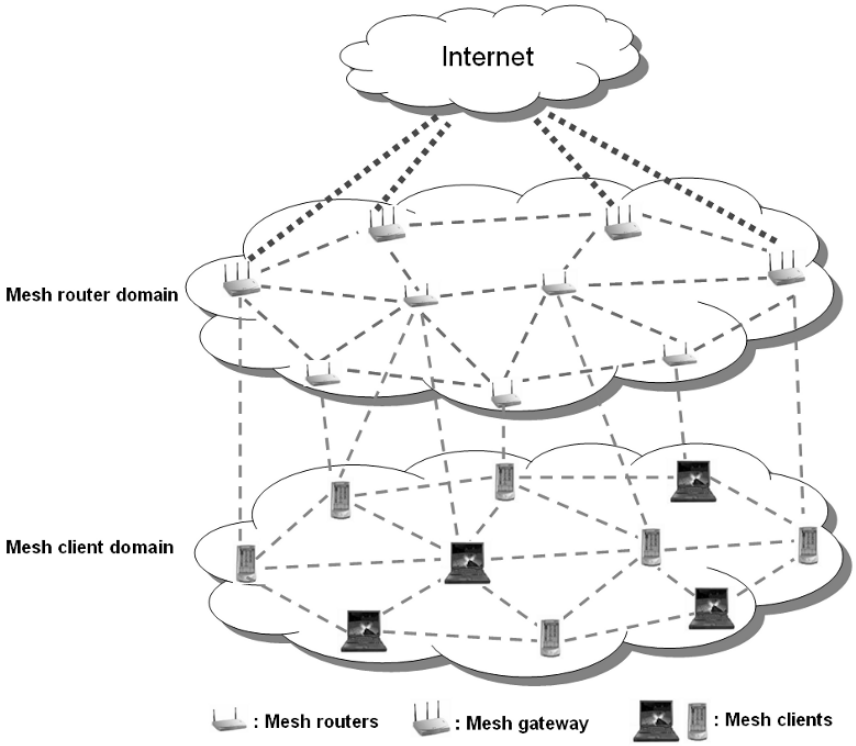



**Fig. 1.1.** An illustration of wireless mesh network architecture. Mesh routers are resource-rich nodes equipped with high processing and memory capabilities, while mesh clients have limited memory and computational power.

## 1.2 Network Architecture

A typical wireless mesh network consists of mesh routers and mesh clients as shown in Fig. 1.1. In this architecture, while static mesh routers form the wireless backbone, mesh clients access the network through mesh routers as well as directly meshing with each other. Unlike a traditional ad hoc network, which is an isolated self-configured wireless network, the mesh network architecture introduces a hierarchy with the implementation of dedicated and power enabled mesh routers. In this integrated network architecture, some of the mesh routers are also called as gateways, which are special wireless routers with a high-bandwidth wired connection to the

Internet. More specifically, mesh routers contain advanced routing functionalities to support mesh networking. This feature of mesh routers is realistic, since mesh routers are fixed nodes, with no constraints on power supply (since they are assumed to be connected to power lines), with multiple wireless interfaces built on either the same or different wireless access technologies.

Different from mesh routers, mesh clients can be mobile nodes, which typically run on batteries. Thus, power usage of mesh clients should be limited. This can be achieved by means of reduced radio functions, e.g., single wireless interface, low antenna gain, and low computational complexity. The target technology for both mesh routers and mesh clients is the IEEE 802.11, which is a well known standard for Wireless Local Area Networks (WLANs). The main reason is the widespread availability of 802.11 devices, which allows a fast deployment of WMNs by using off the shelf solutions[5]. However, to leverage on this opportunity, the modifications required by mesh routers and mesh clients should be aware of the existing hardware constraints and limitations.

### 1.2.1 Open Research Issues

Although wireless mesh networks bring several advantages compared to traditional wireless networks, there exist several open research issues that need to be investigated for the network architecture:

- **Network Planning:** In WMNs, a careful planning of hardware resources needs to be devised in terms of the position and the number of wireless interfaces, and the technology limitations. In fact, a trade-off exists between the number of routers/interfaces, i.e., network cost, and the overall network performance, i.e., network capacity and reliability. Also, how to best place mesh routers impacts the network capacity and topology, and thus, needs to be investigated.
- **Network Provisioning:** A sophisticated network management tool needs to be developed for both mesh routers and mesh clients to dynamically establish connections between them and to closely follow the dynamics of traffic load and users mobility.
- **Network Integration:** It is necessary to design a low-cost method to integrate the IEEE 802.11 and IEEE 802.15.4, and IEEE 802.16 technologies so that a mesh router of one technology can have additional interfaces of another technology. However, which mesh router shall have additional interfaces is also part of the network planning issue.

## 1.3 Design Challenges

The unique characteristics of wireless mesh networks (WMNs) bring many open research issues to the network architecture design and the communication protocols of

[5]In addition to IEEE 802.11 standard, other industrial standard groups, such as IEEE 802.15, and IEEE 802.16, are also actively working on new specifications of wireless mesh networks (see Section 1.6).

WMNs, ranging from the application layer to the physical layer. Although there exist recent advances in the mesh networking technology, many research problems still need to be resolved: the protocols in all communication layers need to be improved, new algorithms are required for efficient network self-configuration, and the network security needs to be ensured. The critical factors influencing the performance of WMNs can be summarized as follows:

- **Advanced Wireless Radio Technologies:** Recently, many solutions have been proposed to improve the capacity of WMNs. Typical examples, include reconfigurable radios, frequency agile/cognitive radios, directional and smart antennas, multiple input multiple output (MIMO) systems, and multi-radio and multi-channel systems. However, the complexity and the cost of these technologies are still too high to be widely accepted for the commercialization. Therefore, all these advanced wireless radio technologies require a revolutionary design in the communication protocol suite in order to facilitate the deployment of WMNs and the commercialization of the products.
- **Interoperability and Integration of Heterogeneous Networks:** Existing networking technologies have limited capabilities of integrating different wireless networks. Thus, to increase the performance of WMNs and to provide the interoperability between the products from different manufacturers, the integration capabilities of multiple wireless interfaces and the corresponding gateway/bridge functions of network routers should be improved.
- **Network Security:** Denial of service attacks and intrusions in WMNs can cause severe damage to the operation of the deployed network. Although there exist many security schemes proposed for wireless local area networks and ad hoc networks, most of these security solutions are either not practical or showing poor performance in WMNs because of the lack of a centralized trusted authority to distribute a public key in the WMN architecture. Consequently, there is a need for new security schemes ranging from efficient encryption and authentication mechanisms to secure key distributions, and intrusion detection mechanisms.
- **Scalability:** The deployed mesh network must be able to deal with large network topologies without increasing the number of network operations exponentially. In addition, the network performance should not degrade as the number of hops between the sender and the receiver increases. To provide the scalability in WMNs, there is a need for scalable MAC, routing and transport layer protocols with minimum overhead.
- **Heterogeneous Quality of Service (QoS) Requirements:** The network services that are provided by WMNs vary from reliable file transfer to real-time multimedia, such as live video streaming. Thus, in addition to traditional network throughput and communication latency metrics, more comprehensive performance metrics, such as delay jitter, aggregate and per-node fairness, and packet loss ratios, need to be considered by the developed mechanisms.
- **Dynamic Network Connectivity and Self-Configuration:** In WMNs, to eliminate the single point failures and potential bottleneck links, the wireless back-

bone needs to provide redundant paths between the sender and the receiver, i.e. mesh connectivity. However, the topology and connectivity of the network can vary frequently because of the route failures and energy depletions[6]. Therefore, to take all the advantages of autonomous mesh connectivity, efficient network self-configuration, topology control and power management algorithms are required.

- **Mobility Support:** To support mobile mesh clients in WMNs, it is necessary to design advanced physical layer and networking techniques, which adapt to the fast fading conditions commonly associated with the mobile users. In addition to these advanced techniques, low latency handover and location management algorithms are also required to improve the quality of service during mobility.
- **Network Management Tools:** To monitor the overall network performance and maintain the network operation, flexible and scalable network management capabilities are required for WMNs. The primary network management capabilities of the WMNs include: i) bandwidth provisioning, ii) installing security and quality of service policies, iii) supporting service level agreements, iv) fault identification and resolution, v) addition and removal of network entities, vi) change of network functions, vii) accounting, billing and reporting. All these capabilities can automate the fault-management in WMNs and thus enable the rapid deployment of WMNs.

## 1.4 Layered Communication Protocols

In this section, we describe the protocol stack of wireless mesh networks and emphasize the open research issues at each communication layer.

### 1.4.1 Application Layer

The necessity to deploy WMNs is determined by the real-world application requirements. Recently, several commercially interesting applications for broadband wireless services have been deployed based on the wireless mesh network architecture. However, since numerous applications can be supported by the WMNs, it is infeasible to have a complete list of them. Here, depending on the functions for WMNs, we categorize the applications of WMNs into several classes:

- **Internet Access:** Recently, several Internet Service Providers (ISPs) deploy wireless mesh networks (WMNs) to enable broadband wireless services in urban, suburban, and rural environments [1] and [16]. These WMN deployments bring significant advantages over traditional wireless networks, including extended network coverage, high speed, and cost-effective network installation. Therefore, the deployments of WMNs are also expected to grow with the increase in demand for broadband wireless Internet access.

[6]Note that in WMNs mesh routers do not have a constraint on power consumption, but the mesh clients usually have limited power resources.

- **Public Safety:** Wireless mesh networks appear to be one of the most promising solutions to address the needs of law enforcement agencies and city governments, such as the police, fire departments, first responders, and emergency services. Currently, several mesh networks are operating to provide mobility support, reliability, flexibility, and high bandwidth for public safety applications [59], [54], [19] and [7]. However, the recent field trials and experiments with existing communication technologies show that the performance of WMNs is still below what they are expected to be. Consequently, there is a need for the development of large-scale physical test-beds and novel communication protocol suites for WMNs.
- **Building Automation:** In a building, the operation of various electrical devices, including ventilation and air conditioning (HVAC) systems, power, light, elevator, etc., needs to be controlled and monitored in real-time. Traditionally, all these operations are realized using wired networks, which is very expensive due to the installation and maintenance costs. In this context, wireless mesh networks can offer efficient and cost-effective solutions for advanced building automation systems.
- **Electric Utility Automation:** In today's competitive electric utility marketplace, electric utilities continuously encounter the challenge of providing reliable power to the end users at competitive prices. Equipment failures, lightning strikes, accidents, and natural catastrophes all cause power disturbances and outages and often result in long service interruptions [24]. WMNs can provide an economically feasible solution for the wide deployment of high speed wireless communications for electric utility automation applications, such as real-time grid and equipment monitoring, incipient fault detection and identification, and wireless automatic meter reading.
- **Information Sharing within the WMNs:** Currently, there exist several peer-to-peer (P2P) networking protocols for information sharing on the Internet. However, the performance of all these P2P protocols may not be high in WMNs, since WMNs have different and unique characteristics compared to the Internet. Therefore, in order to support P2P applications, new well designed protocols need to be integrated into the application layer.
- **Transportation Systems:** Recently, various public transportation companies and the government agencies are interested in practical networking solutions to realize the information delivery system controlling several transportation services [44]. In this regard, wireless mesh networks (WMNs) can provide flexible wireless networking solutions to intelligent transportation systems. With the use of WMNs, the problems of transportation congestion can be addressed and transportation security and safety can be improved.

### Open Research Issues

The main research directions in the WMNs application layer can be classified as follows:

- **Cross-Layer Approach:** To provide strict quality of service (QoS) requirements of the applications and to create application protocols for managing distributed information sharing in WMNs, the protocols in the lower layers need to work interactively with the application layer. This requires a cross-layer approach through information sharing among application, transport, routing, medium access control (MAC) and physical layers. In this way, the deployed WMN can be self-adaptive to network dynamics and meet end-to-end real-time deadlines of the applications.
- **Design of New Applications:** To enable large-scale WMNs and to realize fully integrated and cooperative wireless networking solution, new and commercially interesting applications need to be studied based on the exclusive features and advantages of the WMNs. In this way, this new technology can be made very attractive for both consumers and service providers.
- **Integration of Private and Public Networks:** Novel application protocols that incorporate the use of pricing as an incentive mechanism to encourage private and self-interested nodes to participate in a public wireless mesh network need to be studied. For example, the Internet access can be considered as a service, and hence access points are the service sellers. In this respect, any downstream wireless mesh nodes may purchase this service, for its own consumption, or for reselling it to other downstream nodes obtaining a fair revenue.

### 1.4.2 Transport Layer

To the best of our knowledge, no transport protocol has been introduced specifically for WMNs to date, although several transport protocols have been developed for both wired and wireless networks in the last decade [3]. In this section, we explain existing transport layer protocols with a focus on ad hoc networks, since WMNs share common features with ad hoc networks in spite of their differences. In any case, it is useful to keep in mind that efficient transport protocols are needed for non-real-time and real-time traffic for satisfying different QoS requirements in WMNs.

#### TCP-Based Solutions

Most of the wireless transport protocols proposed in the literature or in use today are enhancements of TCP, which is originally designed to work in the wired Internet [12]. However, TCP-based solutions suffer from various limitations in WMNs due to some inherent properties of TCP and the unique communication challenges of WMNs. The shortcomings of TCP in wireless ad hoc networks have been investigated in [5], [23], [27], [39], [42], [17]. In this section, we briefly discuss some of the significant drawbacks of TCP-based solutions in the context of wireless mesh networks. More specifically, we categorize the discussion based on the following characteristics of TCP-based solutions: (i) under-utilization of network resources and (ii) imprecise congestion detection and control.

- **Under-utilization of Network Resources:** In WMNs, The route failures and consequent route changes affect the congestion control performance of TCP-based solutions significantly. Whenever a route changes, a TCP-based solution employs a slow start mechanism to probe for the available throughput capacity. This mechanism does not allow to increase the rate aggressively, since every connection takes several round-trip-time (RTT) periods before reaching its effective bandwidth value, spending a considerable portion of its lifetime in the probe state. This behavior leads to an under-utilization of network resources, especially for dynamic wireless networks. In addition, the $RTT$ dependence of TCP-based solutions can also be shown through the following analysis. Based on the well-known square root formula [41], when TCP losses occur primarily due to link errors, the TCP throughput of each connection can be represented as a function of $p$ and *RTT*:

$$\Upsilon(RTT, p) \sim \frac{8 \times \nu \times MSS}{RTT \times \sqrt{p}} \tag{1.1}$$

  where $p$ is the error rate, MSS represents the packet length (in bytes), *RTT* is the round trip time delay, and $\nu$ is an implementation specific constant, e.g., $\nu$ is $\sim \sqrt{2}$ for TCP without delayed acknowledgments and $\sim 1$ with delayed acknowledgments. This formula also shows another inefficiency of TCP-based solutions, i.e., the throughput of a TCP flow decreases, when the *RTT* of a connection increases.
- **Imprecise Congestion Detection and Control:** In WMNs, end-to-end congestion detection and control can be imprecise because of the inaccurate estimations of the *RTT* and the dynamic nature of the wireless channel. In multi-hop wireless mesh networks, link failures are frequent and happen either due to the nodes moving out of range of each other, or due to heavy contention, which is perceived as a link breakage on repeated failures to deliver a packet. These breakages lead to route failures, which then result in frequent route re-computations. As different routes may have different round trip times (*RTT*s), measurements of *RTT* on different routes result in large variance in its estimate, leading to large *RTO* (retransmission timeout) values according to the well known formula $RTO = (RTT_{avg}) + 4 \times (RTT_{dev})$ in which $RTT_{avg}$ is the exponentially average of the *RTT* samples observed and $RTT_{dev}$ is the standard deviation of the *RTT* samples.

Based on the TCP drawbacks, which are revealed through many performance evaluation studies, several transport layer solutions have been proposed in the literature for wireless ad hoc networks. All these solutions propose to solve the problems by improving TCP with additional functionalities, modifications, or getting support from lower layers. In this section, we list various enhanced TCP protocols by addressing the proposed solutions to the classical TCP problems on wireless networks. In [23], link level protection and ACKing mechanism were advocated to improve the TCP performance over wireless ad hoc networks. In [5], the problems of TCP in dynamic multihop wireless networks were determined and additional mechanisms at media access and routing layers were proposed to improve TCP performance. The

explicit link failure notification (ELFN) technique was studied in [27], which is based on explicitly informing the TCP source of the link failures to improve TCP performance. In [39], a transport layer solution (ATCP) was proposed, which introduces a thin layer between the transport and underlying routing layers to improve TCP performance by putting TCP into persist mode whenever the network gets disconnected or there are packet losses due to high bit error rate. In [42], a fractional window increment scheme for TCP (TCP-FEW) was proposed to prevent unnecessary network contention by limiting the growth rate of TCP's congestion window. In [17], an adaptive pacing mechanism (TCP-AP) was developed for wireless multi-hop networks in order to avoid bursty packet transmissions.

It is important to note that all these protocols are based on end-to-end rate adjustment and congestion control mechanisms and require a fine-grained end-to-end communication between the source and the destination. Therefore, they may experience significant network inefficiency in WMNs due to the dynamic characteristics of multi-hop wireless environments, end-to-end delay and even obsolete receiver rate feedbacks.

### Novel Transport Protocols

Researchers have developed entirely novel transport protocols for both ad hoc and mesh networks to address the fundamental problems existing in TCP, as argued in the previous sections. In [55], the ad hoc transport protocol (ATP) was proposed for mobile ad hoc networks. The ATP utilizies a rate-based transmission mechanism for rate estimation, and a quick start algorithm for the initial bandwidth estimation. Also, it decouples the congestion related and non-congestion related losses. In this way, the ATP achieves higher performance compared the TCP variants in terms of communication delay, network throughput, and fairness.

Recently, an adaptive and responsive transport protocol (AR-TP) for WMNs has been proposed in [26] in order to fairly allocate the network resources among multiple flows, while minimizing the performance overhead. AR-TP includes both efficient hop-by-hop rate adjustment and reliability mechanisms to achieve high performance reliable data transport in WMNs. Compared to end-to-end rate control schemes, hop-by-hop rate adaptation strategy of the AR-TP protocol enables each router to keep track of dynamic wireless channel conditions in a responsive manner. In addition, with the use of hop-by-hop strategy, the AR-TP can adapt its data transmission rate opportunistically in case of multi-channel WMNs. Performance evaluation via extensive simulation experiments show that the AR-TP protocol achieves high performance in terms of network throughput and fairness.

### Transport Protocols for Real-Time Communication

In WMNs, a real-time rate control protocol is necessary to meet the end-to-end deadlines of the applications [3]. In [20], an adaptive detection rate control (ADTFRC) scheme was developed for wireless ad hoc networks. This protocol proposes a multi-metric joint detection mechanism for TCP-friendly rate control algorithms. However,

the performance of the detection mechanism is not satisfactory to deliver real-time multimedia traffic.

Another end-to-end TCP-friendly rate control protocol for mobile media streaming called RCM was proposed in [57]. RCM does not distinguish between congestion loss and link loss. Once a loss is detected, RCM reduces the sending rate. This behavior makes the protocol power efficient since reducing the sending rate in burst error state may reduce the link corruption. When there is no loss, a new rate increase mechanism, which takes burstiness of loss into account heuristically, was used. Specifically, if a heavier burst loss is detected, a more aggressive increase is applied in order to recover quickly after large rate reduction. However, the existing TCP-friendly rate control protocols [57], [20] cannot be used in WMNs to support real-time delivery for multimedia traffic since all non-congestion packet losses caused by different reasons are handled in the same way. This may degrade the performance of these schemes.

An analytical framework was proposed in [51] for evaluating the quality of service (QoS) of TCP-Friendly Rate Control protocol (TFRC) in hybrid wireless/wired networks. The authors considered a wireless network with the link-level truncated ARQ scheme and limited interface buffer size. They developed one Discrete Time Markov Chain (DTMC) to investigate the wireless bandwidth utilization and packet loss rate of TFRC flows over wireless links, and another DTMC to study the delay outage probability, and the probability of packet delay exceeding a prescribed threshold. Even though extensive simulations were conducted to verify the analytical results, this model may not be applicable to WMNs, since it considers only single hop wireless communication.

In summary, neither specific RCP scheme has been proposed for WMNs nor previous schemes, designed for generic mobile ad hoc networks, have been successfully adapted to the unique features of the WMNs. Therefore, an efficient real-time rate control mechanism for WMNs is still an attractive research area.

### Open Research Issues

To design a reliable and effective transport layer for WMNs, several other issues need to be investigated:

- **Adaptive Transport Protocol Design:** Due to the natural WMNs' integration with many different wired and wireless networks, such as Internet, IEEE 802.11, 802.15, 802.16, etc., the same TCP protocol, designed for a specific network, will be ineffective for the integrated WMNs. On the other hand, using different TCP variants in different wireless networks is not practical. Therefore, the design of a dynamic adaptive transport protocol can be one of the promising transport layer solution for WMNs. Furthermore, new loss differentiation schemes and real-time rate control mechanisms should be developed for multimedia applications in WMNs.
- **Cross-Layer Design:** Since the performance of the transport layer significantly depends on the lower layers, optimizing only transport layer functionalities is insufficient to obtain high performance at the transport layer. Therefore, transport

layer protocols should be jointly optimized with the lower layers by exploiting the tight coupling between each communication layer.

### 1.4.3 Routing Layer

The routing layer is one of the key communication protocol layers to efficiently use the resources in a wireless mesh network, where the available bandwidth is cut down by both internal and external radio interference. The design of a routing algorithm for WMNs should consider the following requirements:

- **Distributed:** The routing algorithm must be distributed, as it is impractical to have a centralized entity computing routes for all the routers. Thus, each router must be able to autonomously take forwarding decisions for every packet.
- **Independent of Any Traffic Profile:** It is not always possible to have an a priori knowledge of the offered traffic load. Therefore, the routing algorithm should not require such a knowledge and should perform well under any traffic profile.

Besides the above requirements, a routing algorithm designed for WMNs should accomplish the following objectives deriving from the unique characteristics of a multi-hop wireless network:

- **Link Quality Variations:** In a wireless mesh network, the quality of a wireless link can rapidly change because of varying environment conditions. The routing algorithm must be able to cope with such changes in link quality and rapidly provide an alternative route in case a link becomes unusable.
- **Reduced Overhead:** Information that the mesh routers exchange as a support to their routing decisions represents an overhead that should be minimized. Indeed, not only it consumes bandwidth on the link it is transmitted over, but it also prevents nodes in the neighborhood from transmitting data.

Furthermore, a multi-radio wireless mesh environment poses additional challenges. In fact, the selection of a neighbor node as the next hop for a packet must take into account which channel is used to communicate with that neighbor node. Selecting a channel, which is being massively used by neighboring nodes to transmit packets, may delay the transmission of the packet and decrease the throughput. Therefore, an appropriate routing strategy must be devised, which takes into account the load on the available wireless channels to reduce the interference and increase the throughput.

In the literature, there exist several routing layer protocols for multi-hop wireless networks [3]. In the next section, we present an overview on the existing routing algorithms for multi-hop wireless mesh networks along with their shortcomings.

#### Multi-Radio Routing

There are a few proposals dealing with routing in multi-radio WMNs. In [48], an iterative algorithm which aims at assigning channels to radio and routing a predefined

traffic profile is presented. However, no new routing algorithm was proposed, as the traffic profile is routed using either the minimum-hop path routing or a randomized multi-path routing. In [56] it was assumed that the set of connection requests to be routed is known. Both an optimal algorithm based on solving a Linear Programming (LP) and a simple heuristic were proposed to route such requests. In [47], distributed channel assignment and routing algorithms were developed. At any time each node joins a gateway node and sends all the packets destined for the wired network to that gateway. Nodes also advertise their cost to reach the gateway they are currently associated with. Cost dynamically changes as it depends on residual bandwidth to achieve load balancing. If a node is notified of a less cost path towards another gateway, it starts a procedure to associate with that gateway. However, such procedure involves updating the routing tables of all the nodes along the paths to the previous and the new gateways. Since cost is dynamic, the proposed strategy may lead to route flaps and a non-convergent network behavior, thus requiring appropriate countermeasures.

**Routing with Various Performance Metrics**

The effect of performance metrics on a routing protocol in static multi-hop wireless Networks was studied in [13]. In this regard, an empirical study was conducted to evaluate the performance of three link-quality metrics, namely ETX (Expected Transmission Count), per-hop RTT (Round-Trip-Time) and per-hop packet pair. All these metrics were studied using a DSR-based routing protocol running in a wireless testbed. In [14], the authors introduced a new metric for routing in multi-radio multi-hop wireless networks, denoted as WCETT (Weighted Cumulative Expected Transmission Time). Such metric explicitly accounts for the interference among links that use the same channel. In [37], a link metric called normalized advance (NADV) was proposed for geographic routing in multihop wireless networks. NADV selects neighbors with the optimal trade-off between proximity and link cost.

**Joint Channel Assignment, Routing and Scheduling**

The joint channel assignment, routing and scheduling problem was investigated in [2] and [34]. In both papers, it was assumed that the knowledge of the traffic demands is available and that the system operates synchronously in a time slotted mode. In [2], an LP was formulated to route the given traffic demands in order to maximize the system throughput subject to fairness constraints. Constraints on the number of radios and on the sum of the flow rates for the links in the interference range were also included. Since the resulting formulation includes integer variables which make the problem NP-hard, the authors solved the LP relaxation of the problem. In [34], the traffic demands were formulated as a multi-commodity flow problem, where one among several different objectives can be defined. Besides including the additional constraints as in [2], the LP formulation in [34] made use of time-indexed variables, hence solving such LP gave a solution for the entire channel assignment, routing and scheduling. However, the presence of integer variables makes the problem NP-hard and thus, the authors solved the LP relaxation of the problem. Then, a channel

assignment along with scheduling based on greedy coloring was used to resolve the potential conflicts.

#### Load Balanced Routing

Load balanced routing in wireless networks has been the subject of many works [21, 31, 36], which however mostly consider a single-radio environment. In [28], a multi-radio scenario was considered, but some simplifying assumptions were made such as modeling the wireless connections between neighbors as isolated point-to-point links. This was achieved by requiring on each node as many radio interfaces as the number of neighbors.

#### Open Research Issues

In WMNs, there exist several open research issues for the routing layer that deserve further investigation:

- **New Link Metrics:** Some link metrics have already been proposed. However, new link metrics may need to be devised to take into account the peculiarities of multi-channel multi-radio wireless mesh networks.
- **Integrated Routing/MAC Design:** In WMNs, the routing layer needs to work interactively with the MAC layer in order to maximize its performance. Integrating adaptive performance metrics from layer-2 into routing protocols or merging certain operations of MAC and routing protocols can be promising approaches.
- **Optimal Flow Value Computation:** In WMNs, the optimal flow distribution can be obtained as a solution to the maximum multi-commodity flow problem, which is NP-complete. An approximation scheme that efficiently fits the routing paradigm should be investigated.

### 1.4.4 Medium Access Control Layer

Wireless mesh networks, being multi-hop networks, are particularly affected by environmental noise and interference problems, as both adjacent hops on the same path and neighboring paths can cause interference. Interference can be alleviated if different node pairs in a neighborhood use non-interfering frequency channels. In case network nodes are equipped with a single radio, the use of multiple channels in the network leads to disconnected subsets of nodes, as each node can only communicate with the neighbor nodes using the same channel. To provide connectivity, new MAC protocols were developed, in [40, 53], which enable nodes to switch their radio to a different channel when needed. However, such an approach presents some drawbacks for WMNs:

- **Synchronization:** The channel switching requires fine-grained synchronization among nodes in order to avoid the *deafness problem*, i.e., the transmitter and the intended receiver may be on different channels.

- **Wasted Time:** The time for channel switching can be in the range of a few milliseconds to a few hundred microseconds [35], which may be unacceptable for most real-time multimedia applications.

Recently, given the availability of low cost wireless devices, another solution to the problem of reducing the interference was proposed, which endows each node with multiple radios. Each radio is set to a different channel and no channel switching is required. Thus, each node can simultaneously communicate over different channels, which was shown to reduce the interference and increase the network throughput [47,48]. However, the limited availability of radios per node and non-overlapped channels requires an efficient assignment of channels to radios. Such an assignment has to satisfy two opposing objectives:

- **Preserve Network Connectivity:** Two neighbor nodes can communicate with each other only if their radio interfaces share a common channel. Thus, the channel assignment must ensure that each mesh router can still communicate (through multiple hops, if it is the case) with all the other routers.
- **Limit Channel Usage:** At the same time, the reuse of the same channel in a neighborhood must be limited, as simultaneous transmissions over the same channel collide, leading to a decrease of the throughput.

In the related literature, there exist several channel assignment algorithms in multi-radio WMNs. In [14], multiple radios per node were used with an *identical* channel assignment, i.e., the first radio is assigned channel 1, the second radio is assigned channel 2 and so on. Such an approach clearly preserves connectivity, but does not make any effort to reduce interference. In [35], a hybrid channel assignment scheme was proposed where some radios are statically assigned a channel while the remaining radios can dynamically change their frequency channel. In [48,56], centralized channel assignment and routing algorithms were introduced. In the proposed channel assignment algorithms the network links are visited in some particular order and a common channel is assigned to the interfaces of both end nodes. If all interfaces of the end nodes in a link are already assigned a channel and they do not share any common channel, then it is necessary to replace one of these channel assignments. Due to the limited number of radios per node, this replacement may trigger a chain reaction and must be performed recursively. The algorithms proposed in [48] and [56] mainly differ in the order in which links are visited and in the criteria used to select the channel to be assigned to a radio. In [48], the channel assignment algorithm visits all the links in decreasing order of the expected link load and selects the channel which minimizes the sum of the expected load from all the links in the interference region that are assigned to the same radio channel. The algorithm proposed by [56] instead visits the links in decreasing order of the number of links falling in the interference range and selects the least used channel in that range.

### Open Research Issues

There exist the following open research issues for the MAC layer:

- **Multi-rate MAC:** A channel assignment algorithm should take into account the availability of multiple physical rates, which presents a challenging trade-off. Indeed, reducing the physical rate decreases the capacity of the link, but also decreases the interference range, thus potentially allowing more simultaneous transmissions.
- **Network Integration:** In WMNs, mesh routers can operate in various wireless technologies, such as IEEE 802.11 and IEEE 802.15.4, and IEEE 802.16. Hence, in the MAC layer, advanced bridging functions should be designed. In this way, different wireless technologies can work together seamlessly [3]. Cognitive and reconfigurable/software radios are one of the promising solutions to these bridging functions.
- **Adaptivity to Network Configuration Change:** In WMNs, new nodes can be joined and some nodes can be left from the network dynamically. Hence, the MAC layer and the associated channel assignment schemes need to be adaptive to these network configuration changes.

### 1.4.5 Physical Layer

Recently, various high-speed physical layer techniques have been developed to improve the capacity of wireless mesh networks. Typical examples include multiple radio interfaces, multiple-input multiple-output (MIMO) systems, beamforming antennas, reconfigurable radios, and frequency agile/cognitive radios. These physical layer techniques enable frequency diversity and multiple transmission rates by a combination of adaptive modulation and different coding rates and thus, increase the network capacity and error resiliency of the radio transmissions. For example, orthogonal frequency multiple access (OFDM) technique has significantly improved the data transmission rate of IEEE 802.11 from 11 Mbps to 54 Mbps. A much higher data transmission rate can be achieved through ultra-wide band (UWB) communications. However, UWB can only be applied for short-distance communications.

In addition, multiple-input multiple-output (MIMO) systems have been developed in order to further improve the capacity and the reliability of WMNs. Specifically, the MIMO systems exploit antenna diversity and spatial multiplexing, which increase network capacity and mitigate the wireless channel impairment by fading, delay-spread, and co-channel interference. MIMO systems can also have different complexities. In WMNs, the MIMO systems with lower complexity are preferred by mesh clients, while those with higher complexity can be applied to mesh routers.

Furthermore, to cope with radio interference and to enhance the multi-hop communication performance and energy efficiency, smart and directional antennas can be utilized in the wireless backbone [3]. The main idea of using smart antennas in WMNs is to exploit the beamforming capability of the transmit/receive antenna arrays. In this way, an effective antenna pattern can be created at the receiver with high gain in the the desired radio signal direction and low gain in all other directions [9]. Therefore, the exploitation of directional transmissions can provide a high speed wireless backbone and improve the spatial reuse. However, the network cost

is a challenging problem in WMNs. In this regard, extensive field tests and research efforts are required to implement fully adaptive and low-cost smart antenna systems.

To exploit the existing wireless spectrum opportunistically in WMNs, frequency-agile or cognitive radios are being developed. According to the Federal Communications Commission (FCC), approximately 70% of the allocated spectrum is not utilized [18]. Moreover, the time scale of spectrum occupancy vary from milliseconds to hours [3]. Therefore, abundant spectrum is still available for wireless communication. Cognitive radio techniques on a software radio platform are one of the most promising solutions to address the limited available spectrum and the inefficiency in the spectrum usage. This is because software radio platforms enable the programmability of all radio components, such as RF bands, channel access modes, and channel modulations, and hence provide flexibility. Although, there are some physical testbeds available, the software radio platforms need further research and field tests.

**Open Research Issues**

There exist the following major challenging issues in the physical layer:

- **Cross-layer Design:** Higher-layer protocols need to work interactively with the physical layer to optimize the networking functions and to fully utilize the advanced physical layer techniques. This leads to cross-layer network design among physical and networking functionalities. However, the cross-layer approach makes hardware design more expensive and challenging. This motivates the use of low-cost software radio platforms in WMNs.
- **Low Cost MIMO Systems:** The complexity and the cost of multiple-antenna systems should be reduced to be facilitate the commercialization of the MIMO products. In order to achieve higher transmission rate in the large deployment fields, new wideband transmission schemes other than OFDM and UWB are also required.
- **Advanced Cognitive Radio Techniques:** Cognitive radio techniques for WMNs are still in their infancy. To achieve viable frequency planning for WMNs, cross-layer spectrum management functionalities, such as spectrum sensing, spectrum decision, and spectrum mobility, need to be investigated [4]. Thus, extensive field tests and research efforts are required before the cognitive radio techniques are accepted for commercial use in WMNs.

## 1.5 Physical Testbeds and Implementations

Recently, a few experimental testbeds have been implemented in the field of wireless mesh networks (WMNs) providing a good basis for implementing and evaluating new protocols and techniques. The Roofnet [8] is an experimental 802.11b/g mesh network in development at MIT, which provides broadband Internet access to users in Cambridge. There are currently around 40 active nodes on the network. This project focused on the effect of routing protocols, node density, and adaptive transmission

rate mechanisms on the overall network performance. The TAPs project [49] designs a wireless mesh network architecture based on Transit Access Points (TAPs). The TAPs form a wireless mesh backbone via high-performance multiple-input and multiple-output (MIMO) wireless links. The focus of this project is the efficient support of multiple antenna and multiple interface systems through the hardware design of the deployed routers. In the Hyacinth project [47], the researchers developed a multi-channel wireless mesh network (WMN) architecture that can be built using IEEE 802.11 a/b/g or IEEE 802.16a technology. In the Hyacinth project, currently ten nodes are equipped with multiple 802.11 radios. The main design issues of this network architecture are: interface channel assignment and packet routing. Moreover, in [11], routing problems and switching delays were investigated based on experiments on a physical testbed, including 20 nodes working on multi-channel environments.

The Broadband and Wireless Networking Laboratory (BWN-Lab) at Georgia Institute of Technology built a testbed of WMNs, including 15 mesh routers and 80 sensor nodes spread across one floor of a building. In this physical test-bed, the effects of the router placement, mobility, link failures and other research issues on the overall network performance were investigated. The deployed WMN testbed was also integrated with wireless sensor and actor networks and adaptive communication protocols were developed for heterogeneous wireless networks [25,26]. Furthermore, the Quail Ridge wireless mesh network [45] is located in Lake Berryssa, California. In this outdoor environment, since the measurements are not affected by external radio interference and other electronic noises significantly, it is easier to investigate the link quality variations. This testbed has been used for wild life monitoring and support audio and video applications. The WINGs project [22] also enables new wireless network architectures, in which all network nodes can move with minimal effect to the network performance. This project is targeted on a two-tier mobile wireless architecture.

In addition to academic research testbeds, some high-profile companies, such as IBM, Intel, Nokia and Microsoft, conduct extensive field tests on various physical test-beds [3]. Specifically, Microsoft [14] is investigating the routing and MAC layer protocols with multi-radio and multi-channels interfaces on its 20 node testbed. Some other companies are also active in the field of wireless mesh networks, through the deployment of municipal mesh networks in several cities. Typical examples are Strix Systems [54], BelAir [7], Tropos [59] and Firetide [19].

Different from these academic research testbeds and commercial installations, several community wireless mesh networks have also deployed WMNs to study the impact of different technologies, such as multiple channels and directional antennas, and to evaluate the performance of different applications on mesh networks. Few examples of WMN architectures include Wireless Leiden [15], the Digital Gangetic Plains project [46], Seattle Wireless [50], and the TibTec Dharamshala wireless mesh community network [58].

Although all these measurements and experiments provide valuable insight into the advantages of wireless mesh networks, the recent field trials and experiments show that the performance of wireless mesh networks is still below than what they

are expected to be. Consequently, there is a need for the development of large-scale physical test-beds and novel communication protocol suites for WMNs. In addition, international standards are needed for building commercially interesting mesh network applications and customer services on top of WMN architectures.

## 1.6 Standardization Activities

International standards are crucial for the industry since they provide the interoperability between the products from different manufacturers and facilitate the commercialization of the equipments. Depending on the target network type and the application requirements, several standard groups, such as IEEE 802.11, IEEE 802.15, and IEEE 802.16, are actively working on new specifications for WMNs. In the following section, we present the overview of these international standards.

### 1.6.1 IEEE 802.11s Mesh Networks

The initial specifications for the most popular standard for Wireless Local Area Network were completed by the IEEE in 1999 [29] and successively extended in 2003. All the family of IEEE 802.11 standards is specified for one-hop communications, making it unsuitable for multihop, multichannel, and multiradio operations. Hence, the IEEE set up a new working group: 802.11s task group, for the installation, configuration and operation of 802.11-based mesh networks.

In the IEEE 802.11s standard, all the devices, which support mesh functionalities are defined as mesh point (MP). A wireless distribution system (WDS) is a set of MPs and mesh links. In the proposed standard, there are also the mesh access point (MAP), which is a specific MP, but acts as an access point and the mesh portal point (MPP), which is another type of MP through which multiple 802.11-based mesh networks can be interconnected. We can distinguish two different process of initialization in the IEEE 802.11s standard: i) the association of a device with a MAP, which is performed through the usual 802.11 procedure, and ii) the association of MAP with a neighboring node, which is performed after scanning, neighbor discovery, authentication of the MP and channels negotiation.

The main components of the proposed IEEE 802.11s medium access coordination function (MCF) include mesh measurement, mesh interworking, medium access coordination, mesh topology learning, routing and forwarding, topology discovery and association, mesh security, mesh configuration and management, and 802.11 service integration [38]. With these coordination components, the IEEE 802.11s standard addresses the 802.11-based wireless mesh services. It is also important to note that these functionalities can be built on top of the existing physical layer of IEEE 802.11a/b/g/n standards. In the following paragraphs, the major services and requirements of the MCF are explained briefly:

- **Topology Discovery and Association:** When an MP is willing to join the network, it looks for the existing networks, and if no networks are detected, it forms

a new one. The necessary network formation information is gathered either by the passive listening of beacon messages, or by the active sending of probing messages. After the discovery phase, the MPs create the mesh network by associating with the neighboring nodes. It is also possible to create different smaller subnetworks, operating on different channels.

- **Medium Access Coordination:** In [6], [52], the medium access coordination function was proposed based on the enhanced distributed channel access (EDCA) mechanism used in IEEE 802.11e. The proposed MAC algorithms avoid beacon collision, and provide synchronization, congestion control and power saving.
- **Mesh Configuration and Management:** In WMNs, automated management tools are necessary to monitor the overall network performance and maintain the network operation. In this way, the burden of manual configuration for the service provider can be minimized. In addition, mesh networks need to support IEEE 802.11h so that dynamic frequency selection (DFS) requirements can be ensured.
- **Mesh Security:** The functionalities of network security can be based on IEEE 802.11i standard, which specifies the features for security in all WLANs. Moreover, it has to consider high mobility nodes, because they require very frequent authentications.
- **Mesh Interworking with Other Networks:** In case the device is a MPP, it is supposed to offer interconnection with other mesh networks, in order to set up integrated mesh networks.
- **Mesh Measurements:** All the measurements made on the network should be available for the upper layers. In this way, the network capacity and reliability can be improved.
- **Service Integration:** The IEEE 802.11s standard considers the service integration in order to insure full compatibility with other 802.11 networks.

### 1.6.2 IEEE 802.15 Mesh Networks

The IEEE 802.15 Working Group is committed to develop consensus standards for Wireless Personal Area Networks (WPANs) or short distance wireless networks. These WPANs address wireless networking of portable and mobile computing devices, such as PCs, Personal Digital Assistants (PDAs), peripherals, cell phones, pagers, and consumer electronics. They also allow these devices to communicate and interoperate with each another.

To address high data rate solutions in WPANs, IEEE 802.15.3 working group was established and the MAC and PHY layer specifications for high rate WPANs were completed in 2003. This standard targets data rates from 11 to 55 Mbps at distances of greater than 70 m while maintaining quality of service for the data streams. In addition, the IEEE 802.15.4 working group was chartered to investigate a low data rate solution with long battery life and low complexity requirements. A first standard was published in 2003 and then superseded by a new one in 2006. The supported data rates are 250 Kbps, 40 Kbps, and 20 Kbps. The transmission distance is ex-

pected to range from 10 to 75 m, depending on the transmission power output and environmental conditions.

More recently, a new task group, IEEE 802.15.5 started its operation to determine the necessary mechanisms that must be present in the PHY and MAC layers of WPANs to enable mesh networking. Specifically, this task group works to provide an architectural framework for scalable and interoperable wireless mesh topologies for WPAN devices.

Specifically, a mesh WPAN differs from a WPAN in that a direct communication among participating devices in the network is not always possible despite the small covered area. Links in an WPAN can be broken due to the combination of low transmission power and persistent interference or the attenuation by barriers like walls, doors, tables, etc. The mesh infrastructure may cancel these attenuations, which can have a severe effect on the link performance. The mesh WPANs bring significant advantages over a WPANs, including extended network coverage, increased network reliability, and longer network life time. However, the design of extensions for the establishment of mesh WPANs must take the following issues into account:

- **Energy Efficiency:** In WMNs, the power consumption is restricted severely if the device is currently not connected to a power supply. A mesh WPAN MAC should be able to differentiate between power sensible and connected devices and assign different functions to them, so that the power sensible device is able to save power and participate in the network as long as possible.
- **Interoperability:** Non-mesh enabled devices may be present in the same area of the mesh network. Efficient mechanisms are needed that handle the coexistence of those devices with mesh-enabled devices, so that the mesh network is still functional and single-hop transmissions from/to the other non-mesh enabled devices are possible.

**Table 1.1.** Summary of IEEE 802.16 standards.

| Standard | Description |
|---|---|
| IEEE 802.16a | Containing new specifications for the 2-11 GHz bands and the mesh mode |
| IEEE 802.16b | Providing quality-of-service (QoS) feature |
| IEEE 802.16c | System profiles for 10-66 GHz operations and supporting interoperability |
| IEEE 802.16d | Fixed broadband wireless specification |
| IEEE 802.16e | Amendment to the 802.16d specification, explicit support for mobility |
| IEEE 802.16f | 802.16 Management Information Base (MIB) |
| IEEE 802.16g | Providing efficient handover and QoS |
| IEEE 802.16h | Coexistence in license exempt frequency bands |

### 1.6.3 IEEE 802.16 Mesh Networks

To address the broadband wireless access in wireless metropolitan area networks, the IEEE 802.16 working group (WG) was established in 1999. The initial IEEE 802.16 standard[7] was designed to operate in the licensed 10-66 GHz frequency band and to employ a point-to-multipoint (PMP) architecture where each base station (BS) serves a number of subscriber stations (SSs) in the deployment field. However, these operational features require the line-of-sight (LOS) communications, since only a limited amount of multipath interference can be tolerated at the high operating frequencies, i.e., larger than 10 GHz. To address reliable non-LOS (NLOS) operations and to expand the system in non-licensed bands, the IEEE 802.16a extension was ratified in January 2003. The IEEE 802.16a standard operates in a lower frequency of 2-11 GHz, enabling non-line-of-sight communications and meshing functionalities in addition to PMP mode. In Table 1.1, the main specifications of the IEEE 802.16 standard extensions are also summarized.

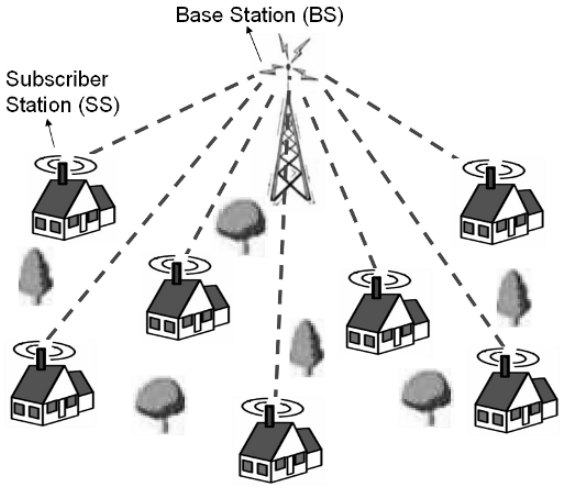
In the IEEE 802.16a standard, the main difference between the PMP mode and the mesh mode is the ability of multihop communication in the mesh mode. As shown in Fig. 1.2, while in the mesh mode SSs can directly communicate with each other through multihop communications, the PMP mode requires each SS to be connected to a central BS through single hop communication. Consequently, the mesh mode enables SSs to relay each others traffic towards the mesh BS, which also connects the SSs to the backhaul network.

Furthermore, in the mesh mode there are two types of TDMA-based packet scheduling mechanisms: centralized scheduling and distributed scheduling. In the centralized scheduling, the BS assigns the radio resources for all SSs within a certain hop range. On the other hand, in the distributed scheduling, all nodes, including the BS, coordinate with each other for accessing the channel. During this coordination, all the nodes broadcast their schedules, i.e., available resources, requests, and grants, to all their neighbors within their two-hop neighborhood.

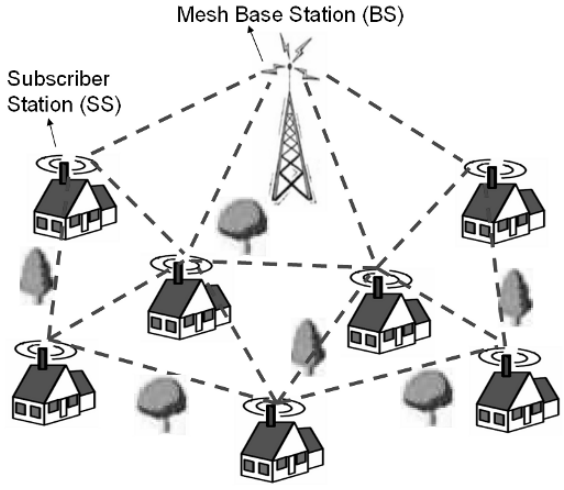
While the IEEE 802.16a mesh mode offers several opportunities, it has also two main drawbacks: i) it addresses only fixed broadband communication applications, and ii) it is not compatible with the existing PMP mode [38]. In order to address these drawbacks, in July 2005, another study group called the Mobile Multihop Relay (MMR) was established under the IEEE 802.16 working group. The main objective of the MMR study group is to support mobile stations (MS) by using multihop relaying techniques using relay stations (RS). An RS relays information between an SS/MS and a BS or between other RSs or between an RS and a BS [43]. Therefore, unlike the mesh mode, the dedicated RSs forms a treelike topology for relaying the traffic to the BS.

Although the main specifications of the IEEE 802.16 standards have been released, few commercial equipments compliant with these international standards have just appeared on the market. To facilitate the deployment of broadband wireless

[7] WiMAX (Worldwide Inter-operability for Microwave Access) is the commercialization of the maturing IEEE 802.16 standard.




(a) Point-to-multipoint (PMP) mode.




(b) Mesh mode.

**Fig. 1.2.** An illustration of the operation modes in the IEEE 802.16a standard.

mesh networks based on the IEEE 802.16 standards and to provide the interoperability between the products from different manufacturers, the WiMAX forum was established [60]. This forum acts like the Wi-Fi and ZigBee Alliances, which are working to promote the IEEE 802.11 and IEEE 802.15.4 standards for WLANs and wireless sensor networks, respectively.

## Conclusion

Wireless Mesh Network (WMN) is a promising wireless technology for several emerging and commercially interesting applications, e.g., broadband home networking, community and neighborhood networks, coordinated network management, intelligent transportation systems. Different from traditional wireless networks, WMN is dynamically self-organized and self-configured. The self-configuration feature of WMNs brings many advantages for the end-users, such as low up-front cost, easy network maintenance, robustness, and reliable service coverage. Although the WMN offers many opportunities, recent field trials and experiments with existing communication technologies show that the performance of WMNs is still below than what they are expected to be. Consequently, there is a need for the development of large-scale physical test-beds and novel communication protocol suites for WMNs. In addition, many open research issues, such as security, quality of service, scalability, distributed network management and self-configuration, optimal network design and configuration, and integration of heterogeneous networks, need to be resolved.

## Acknowledgment

The authors would like to thank Prof. Ian F. Akyildiz for his constructive comments.

## References

1. "The Aiirmesh" - Online Available: http://www.aiirmesh.com/cerritos/
2. M. Alicherry, R. Bhatia, and E. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 1960-1971, November 2006.
3. I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *Computer Networks Journal (Elsevier)*, March 2005.
4. I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks Journal (Elsevier)*, September 2006.
5. V. Anantharaman, S.J. Park, K. Sundaresan, and R. Sivakumar, "TCP performance over mobile ad-hoc networks: A quantitative study," *Journal of Wireless Communications and Mobile Computing*, vol. 4, no. 2, pp. 203-222, 2004.
6. Aoki et al., "802.11 TGs Simple Efficient Extensible Mesh (SEE-Mesh) Proposal," IEEE 802 11-05/0562r01, 2005.
7. "Belair networks" - Online Available: http://www.belairnetworks.com
8. J. Bicket, S. Biswas, D.Aguayo, and R. Morris, "Architecture and evaluation of an unplanned 802.11b mesh network," in *Proc. of ACM MOBICOM '05*, 2005.
9. R. Bruno, M. Conti, and E. Gregori, "Mesh networks: Commodity multihop ad hoc networks", *IEEE Communications Magazine*, vol. 43, no. 3, pp. 123-131, March 2005.
10. "BWN Lab Wireless Mesh Network Test-bed" - Online Available: http://www.ece.gatech.edu/research/labs/bwn/mesh/

11. C. Chereddi, P. Kyasanur, and N. H. Vaidya, "Design and implementation of a multi-channel multi-interface network," in *Proc. of REALMAN Workshop*, 2006.
12. I. Chlamtac, M. Conti, and J. Liu, "Mobile ad-hoc networking: Imperatives and challenges," *Ad Hoc Networks*, vol. 1, no. 1, pp. 13-64, 2003.
13. R. Draves, J. Padhye, and B. Zill, "Comparison of routing metrics for static multi-hop wireless networks," in *Proc. of ACM SIGCOMM '04*, pp. 133-144, 2004.
14. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. of ACM MOBICOM '04*, 2004.
15. R. V. Drunen, J. Koolhaas, H. Schuurmans, and M. Vijn, "Building a wireless community network in the Netherlands," in *Proc. of USENIX/Freenix Conference*, 2003.
16. "Earthlink" - Online Available:http://www.earthlink.net
17. S. M. ElRakabawy, A. Klemm, and C. Lindemann, "TCP with adaptive pacing for multihop wireless networks," in *Proc. of ACM MOBIHOC*, 2005.
18. "FCC Cognitive Radios" -Online Available: http:// www.fcc.gov/oet/cognitiveradio.
19. "Firetide" - Online Available: http://www.firetide.com
20. Z. Fu, X. Meng, and S. Lu, "A transport protocol for supporting multimedia streaming in mobile ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 21, Dec. 2003.
21. J. Gao and L. Zhang, "Load balanced short path routing in wireless networks," in *Proc. of IEEE INFOCOM '04*, vol. 2, pp. 1098-1107, 2004.
22. J. Garcia-Luna-Aceves, C. Fullmer, E. Madruga, D. Beyer, and T. Frivold, "Wireless internet gateways (wings)," in *Proc. of IEEE MILCOM*, 1997.
23. M. Gerla, K. Tang, and R. Bagrodia, "TCP performance in wireless multi hop networks," in *Proc. of IEEE WMSCA*, 1999.
24. V. C. Gungor and F. Lambert, "A survey on communication networks for electric system automation," *Computer Networks Journal (Elsevier)*, vol. 50, pp. 877-897, May 2006.
25. V. C. Gungor, P. Pace, and E. Natalizio, "AR-TP: An adaptive and responsive transport protocol for wireless mesh networks," in *Proc. of IEEE ICC '07*, Glasgow, Scotland, June 2007.
26. V. C. Gungor, C. Sastry, Z. Song, and R. Integlia, "Resource-aware and link-quality-based routing metric for wireless sensor and actor networks," in *Proc. of IEEE ICC '07*, Glasgow, Scotland, June 2007.
27. G. Holland and N. H. Vaidya, "Analysis of TCP performance over mobile ad hoc networks," in *Proc. of ACM MOBICOM '99*, 1999.
28. P. Hsiao, A. Hwang, H. Kung, and D. Vlah, "Load-balancing routing for wireless access networks," in *Proc. of IEEE INFOCOM '01*, vol. 2, pp. 986-995, 2001.
29. IEEE Std 802.11, "Information Technology - Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999 ed. (Rel. 2003).
30. IEEE P802.11e/D13.0, "Amendment: Medium Access Control (MAC) Quality of service (QoS) Enhancements," Jan. 2005.
31. A. Iwata, C. Chiang, G. Pei, M. Gerla, and T. Chen, "Scalable routing strategies for ad-hoc wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1369-1379, August 1999.
32. K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. of ACM MOBICOM '03*, San Diego, CA, 2003.
33. J. Jun and M. L. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Communications Magazine*, October 2003.

34. M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. of ACM MOBICOM '05*, 2005.
35. P. Kyasanur and N. H. Vaidya, "Routing and interface assignment in multi-channel multi-interface wireless networks," in *Proc. of IEEE WCNC '05*, vol. 4, pp. 2051-2056, 2005.
36. S. J. Lee and M. Gerla, "Dynamic load-aware routing in ad hoc networks," in *Proc. of IEEE ICC '01*, vol. 10, pp. 3206-3210, 2001.
37. S. Lee, B. Bhattacharjee, and S. Banerjee, "Efficient geographic routing in multihop wireless networks," in *Proc. of ACM MobiHoc*, 2005.
38. M. J. Lee, J. Zheng, Y. Ko, and D. M. Shrestha, "Emerging standards for wireless mesh technology," *IEEE Wireless Communications*, April 2006.
39. J. Liu and S. Singh, "ATCP: TCP for mobile ad hoc networks," *IEEE Journal on Selected Areas in Communications,* vol. 19, no. 7, pp. 1300-1315, 2001.
40. Y. Liu and E. Knightly, "Opportunistic fair scheduling over multiple wireless channels," in *Proc. of IEEE INFOCOM '03*, vol. 2, pp. 1106-1115, 2003.
41. M. Mathis, J. Semke, J. Mahdavi, and T. Ott, "The macroscopic behavior of the TCP congestion avoidance algorithm," *ACM SIGCOMM Computer Communication Review*, vol. 27, no. 3, pp. 67-82, 1997.
42. K. Nahm, A. Helmy, and C. C. J. Kuo,"TCP over multihop 802.11 networks: Issues and performance enhancement," in *Proc. of ACM MobiHoc*, 2005.
43. Peterson et al., "IEEE 802.16mmr-06/007: Definition of Terminology Used in Mobile Multihop Relay," 2006.
44. "Portsmouth" - Online Available:http:// http://www.portsmouth.gov.uk/
45. D. Wu, D. Gupta, S. Liese, and P. Mohapatra "QuRiNet: Quail ridge reserve wireless mesh network," in *Proc. of the First ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation and Charecterization (WinTECH 2006).*
46. B. Raman and K. Chebrolu, "Revisiting mac design for an 802.11-based mesh network," in *Proc. of HotNets*, 2004.
47. A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. of IEEE INFOCOM '05*, vol. 3, pp. 2223-2234, 2005.
48. A. Raniwala, K. Gopalan, and T. Chiueh, "Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks," *ACM Mobile Computing and Communications Review*, vol. 8, no. 2, pp. 50-65, April 2004.
49. "Rice university Taps Project" - Online Available: http://taps.rice.edu
50. "Seattle wireless" - Online Available: http://www.seattlewireless.net
51. H. Shen, L. Cai, and X. Shen, "Performance analysis of TFRC over wireless link with truncated link-level ARQ," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, June 2006.
52. Sheu et al., "802.11 TGs MAC Enhancement Proposal," IEEE 802 11-05/0575r4, 2005.
53. J. So and N. Vaidya, "Multi-channel MAC for ad hoc networks: Handling multi-channel hidden terminals using a single transceiver," in *Proc. of ACM MobiHoc*, pp. 222-233, 2004.
54. "Strix systems" - Online Available: http://www.strixsystems.com
55. K. Sundaresan et al.,"ATP: A reliable transport protocol for ad-hoc networks," in *Proc. of ACM MobiHoc*, 2003.
56. J. Tang, G. Xue, and W. Zhang, "Interference- aware topology control and QoS routing in multi-channel wireless mesh networks," in *Proc. of ACM MobiHoc*, pp. 68-77, 2005.
57. K. Tan, Q. Zhang, and W. Zhu, "An end-to-end rate control protocol for multimedia streaming in wired-cum-wireless environments," in *Proc. of ISCAS*, vol. 2, May 2003.

58. “Tibetan Technology Center” - Online Available: http://www.tibtec.org
59. “Tropos” - Online Available: http://www.tropos.com
60. “The WiMAX forum” - Online Available: http://www.wimaxforum.org/home/

# 2

# Architectures and Deployment Strategies for Wireless Mesh Networks

J.-H. Huang, L.-C. Wang, and C.-J. Chang

National Chiao-Tung University, Taiwan, R.O.C.
hjh@mail.nctu.edu.tw, {lichun,cjchang}@cc.nctu.edu.tw

## 2.1 Introduction

Nowadays the development of the next-generation wireless systems (e.g., the fourth-generation (4G) mobile cellular systems, IEEE 802.lln, etc.) aims to provide high data rates in excess of 1 Gbps. Thanks to its capability of enhancing coverage with low transmission power, wireless mesh networks (WMNs) play a significant role in supporting ubiquitous broadband access [1]- [10].

Fig. 2.1 illustrates a multi-hop wireless mesh network, where only the central gateway $G$ has a wireline connection to the Internet and other nodes (like node $S$) access to the central gateway via a multi-hop wireless communication. Each node in the WMN should operate not only as a client but also a relay, i.e., forwarding data to and from the Internet-connected central gateway on behalf of other neighboring nodes. The main difference between ad hoc networks and wireless mesh networks is the traffic pattern [2], as shown in Fig. 2.2. In a WMN, there will exist a central gateway and most traffic is either to/from the central gateway as shown in Fig. 2.2(a). In an ad hoc network, however, traffic flows are arbitrary between pairs of nodes, such as the flow between nodes $S1$ and $D1$ in Fig. 2.2(b).

In general, the advantages of wireless mesh networking technology can be summarized into five folds. First, WMN can be rapidly deployed in a large-scale area with a minimal cabling engineering work so as to lower the infrastructure and deployment costs [1]- [5]. Second, mesh networking technology can combat shadowing and severe path loss to extend service coverage area. Third, by means of short range communications, WMN can improve transmission rate and then energy efficiency. In addition, the same frequency channel can be reused spatially by two links at a shorter distance. Fourth, due to multiple paths for each node, an appealing feature of WMNs is its robustness [9], [10]. If some nodes fail (like node $B$ in Fig. 2.3), the mesh network can continue operating by forwarding data traffic via the alternative nodes. Fifth, WMN can concurrently support a variety of wireless radio access technologies, thereby providing the flexibility to integrate different radio access networks [6]- [8]. Fig. 2.4 shows an example of integrated wireless mesh network, where 802.16 (WiMAX), 802.11 (WiFi), and 802.15 (Bluetooth and Zigbee)

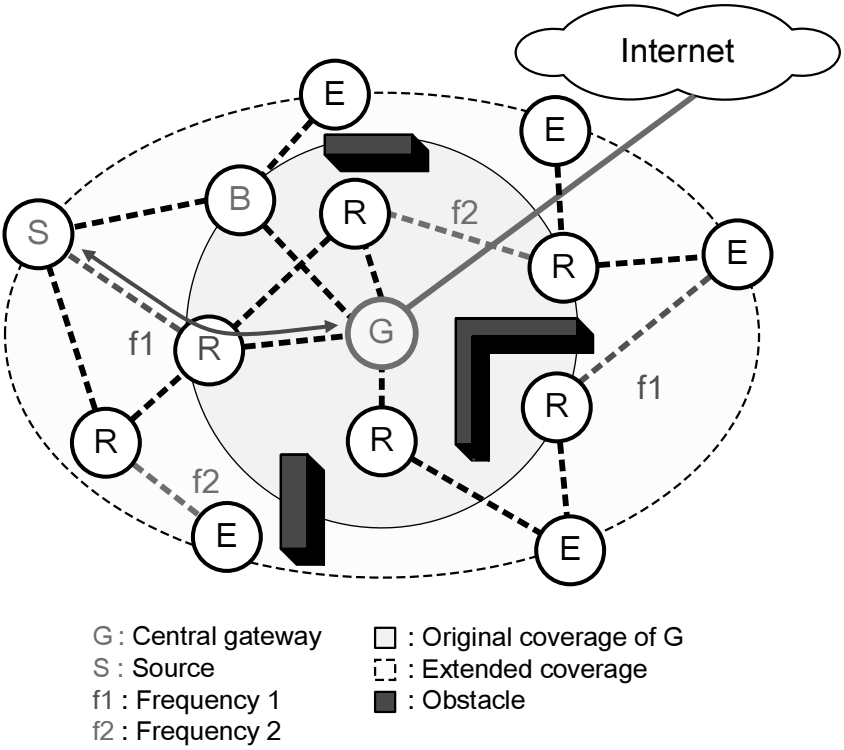
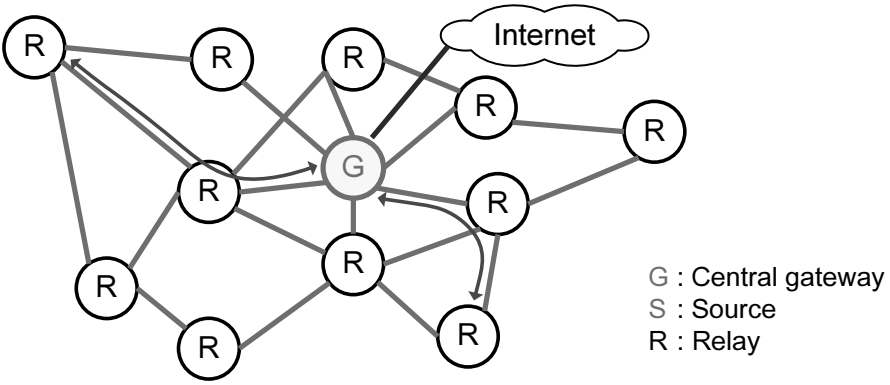



**Fig. 2.1.** Conceptual illustration of a multi-hop wireless mesh network.

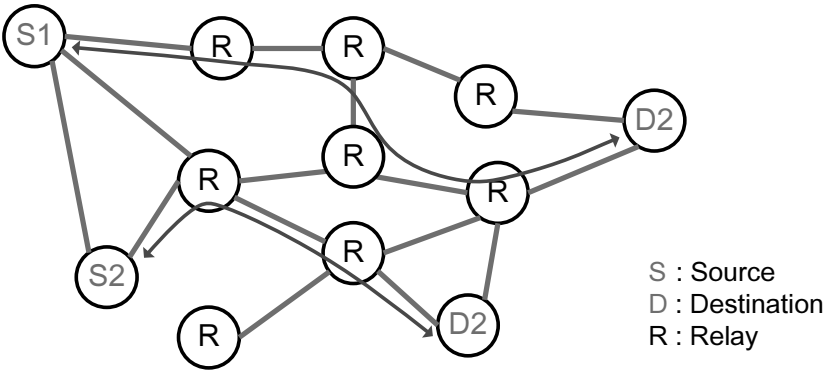
technologies are used for the wireless metropolitan area network (WMAN), the wireless local area network (WLAN), and the wireless personal area network (WPAN), respectively.

However, when the coverage area increases to serve more users, multi-hop networking suffers from the scalability issue [10]. This is because in the multi-hop WMNs throughput enhancement and coverage extension are two contradictory goals. On one hand, the multi-hop communications can extend the coverage area to lower the total infrastructure cost. On the other hand, as the number of hops increases, the repeatedly relayed traffic will exhaust the radio resource. In the meanwhile, the throughput will sharply degrade due to the increase of collisions from a large number of users. Therefore, it becomes an important and challenging issue to design a scalable wireless mesh network, so that the coverage of a WMN can be extended without sacrificing the system overall throughput.

In this chapter, we first discuss the major architectures of WMNs and briefly overview the existing mesh networking technologies, including the IEEE 802.11s and IEEE 802.16 systems. Then, we address the scalability issue of the WMN from a network deployment perspective. We introduce two scalable-WMN deployment strategies for the *dense-urban coverage* and *wide-area coverage scenarios* as shown




(a) Wireless mesh network




(b) Ad hoc network

**Fig. 2.2.** Comparisons of a wireless mesh network and an ad hoc network.

in Figs. 2.5 and 2.6 ( [11, 12]). First, the cluster-based wireless mesh network for the dense-urban area is shown in Fig. 2.5. In this WMN, several adjacent access points (APs) form a cluster and are connected to the Internet through the same switch/router. In each cluster, only the central access point $AP_0$ connects to the Internet through the wires. Other APs are interconnected by wireless links. By doing so, the network deployment in the urban area becomes easier because the cabling engineering work is reduced. Second, a scalable multi-channel ring-based WMN for wide-area coverage is shown in Fig. 2.6, where the central gateway and stationary mesh nodes in the cell form a multi-hop WMN. Note that the mesh cell is divided into several rings allocated with different channels. In the same ring, the mesh nodes can follow the legacy IEEE 802.11 medium access control (MAC) protocol to share the radio

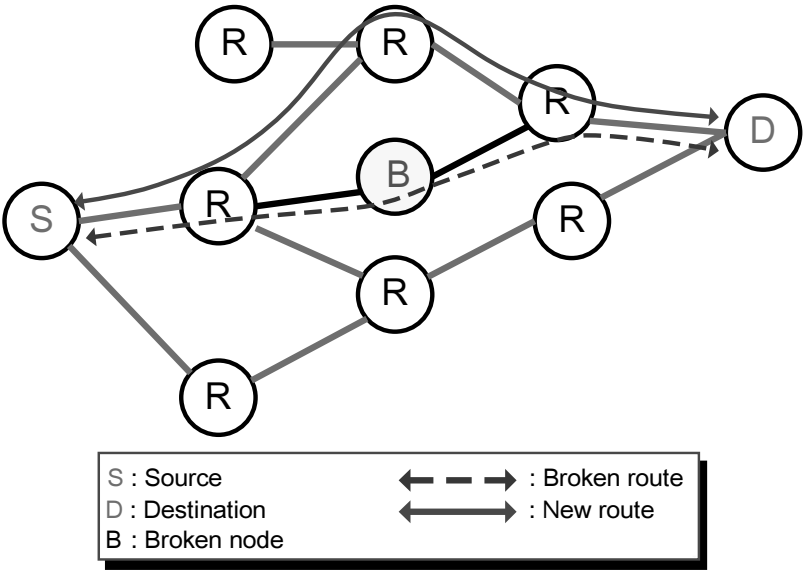



**Fig. 2.3.** Robustness of wireless mesh network.

medium. Besides, mesh nodes in the inner rings will relay data for nodes in the outer rings toward the central gateway. Based on this mesh cell architecture, the service coverage of the central gateway/AP can be effectively extended with a lower cost.

We will also investigate the optimal tradeoff between capacity and coverage for these two scalable WMNs. Most traditional wireless mesh networks are not scalable to the coverage area because the user throughput is not guaranteed due to the increase of collisions. By contrast, the WMNs shown in Figs. 2.5 and 2.6 are more scalable in terms of coverage because frequency planning with multiple channels can be easily applied in this architecture to resolve the contention issue. Thus the throughput can be ensured by properly determining the deployment parameters. We will apply the mixed-integer nonlinear programming (MINLP) optimization approach to determine the optimal deployment parameters, aiming to maximize the capacity and coverage simultaneously.

The rest of this chapter is organized as follows. Section 2.2 presents the major network architectures for WMNs. Sections 2.3 and 2.4 discuss the mesh networking technologies in the IEEE 802.11s and IEEE 802.16 systems, respectively. Section 2.5 describes the proposed scalable wireless mesh networks for the dense-urban coverage and the wide area coverage. In addition, we apply the optimization approach to determine the optimal deployment parameters, aiming at maximizing the coverage and capacity. At last, concluding remarks are given.

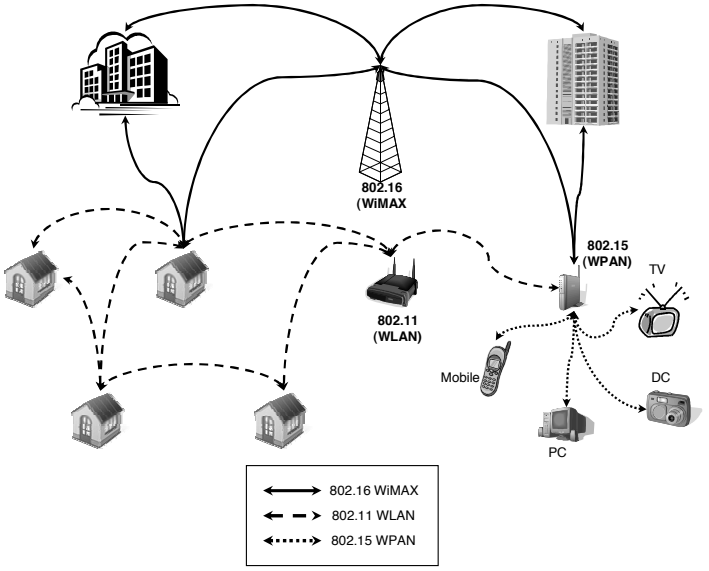



**Fig. 2.4.** An integrated 802.15/11/16 (WPAN/WLAN/WMAN) wireless mesh network.

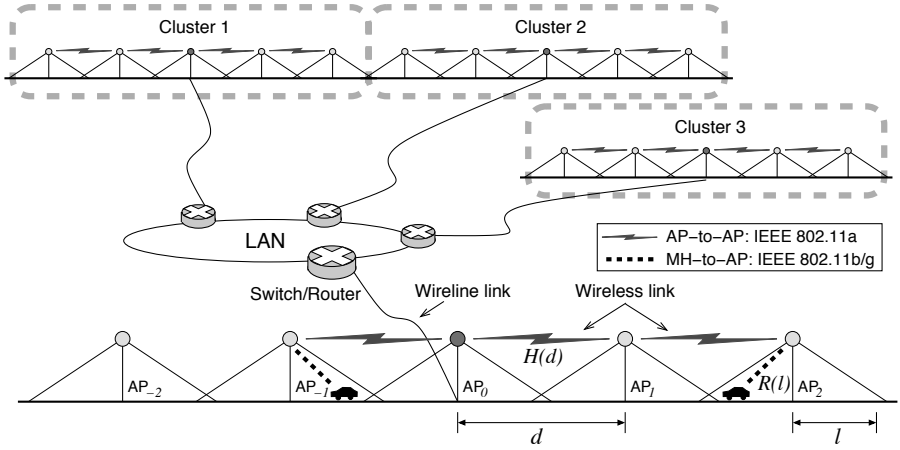



**Fig. 2.5.** Clusters of access points in the wireless mesh network for the dense-urban coverage.

## 2.2 Architectures for Wireless Mesh Networks

A wireless mesh network is an economical and low-power solution to support the ubiquitous broadband services. To provide uniform data-rate coverage, one straightforward solution is to densely deploy base stations (BSs) or access points (APs)

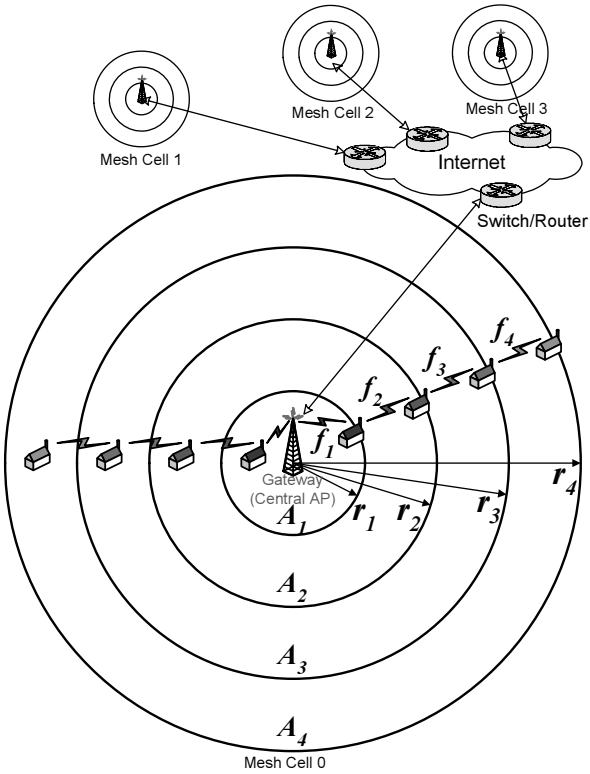



**Fig. 2.6.** Ring-based cell architecture in the wireless mesh network for wide-area coverage, where each ring is allocated with different allocated channel.

in the service area.[1] Fig. 2.7 shows an example of conventional broadband cellular/hotspot network, where all BSs are connected to the Internet via cables. Clearly, such a network architecture is not very feasible due to the high costs of expensive infrastructure and cabling engineering. Recently, mesh networks have become an interesting option for deploying the wireless broadband networks. In the WMN, only the central gateway has wireline connections to access the Internet directly. All the BSs are interconnected via wireless links. By means of low-power multi-hop communications, the coverage can be significantly extended. In addition, deploying such a network is easier owing to less cabling engineering work.

In the following, we discuss the major WMN architectures.

[1]Usually, the term *base station* is used for the traditional cellular systems, while *access point* used for the WLAN-based systems. Unless otherwise indicated, the term base station will refer to both the cellular BS and the WLAN AP.

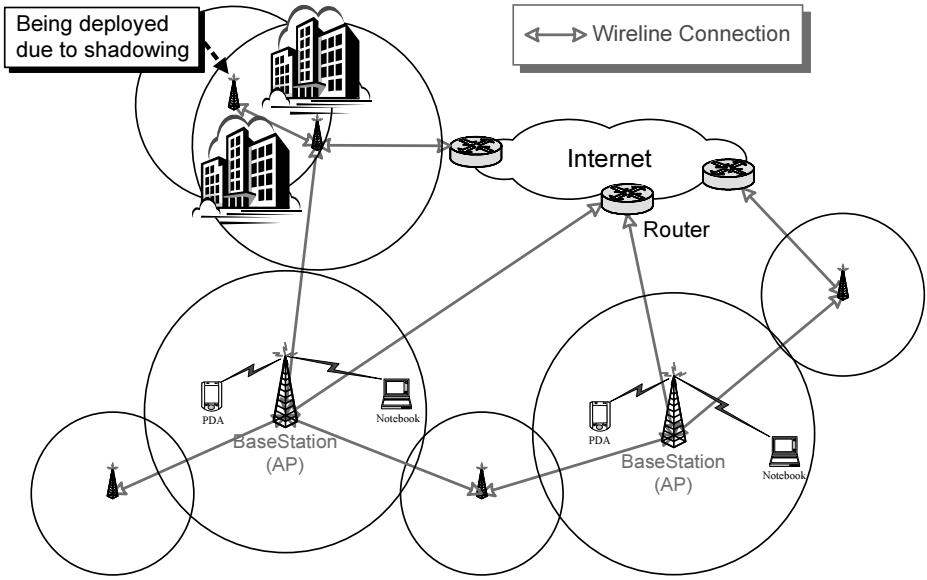



**Fig. 2.7.** Conventional cellular/hotspot broadband network architecture.

### 2.2.1 Backbone Wireless Mesh Network

Fig. 2.8 shows an example of *backbone wireless mesh networks*. In the figure, each base station also operates as a wireless relay to forward neighboring BS's traffic to the gateway. Such a wireless multi-hop backbone network provides the flexibility to integrate WMNs with the existing wireless communication systems. The base stations can concurrently integrate 2G/3G/WLAN/4G radio access technologies to provide voice and high-rate data services, and flexibly employ the emerging broadband radio technologies in the backbone networks.

The backbone WMN has the advantage of incremental deployment [2]. If necessary, more gateways can be added, by simply connecting more base stations to the Internet via wineline. Deploying more gateways in the WMNs can improve not only the network capacity but also the reliability. That is, if one gateway fails, the traffic can be delivered by alternative routes and gateways.

### 2.2.2 Backbone with End-user Wireless Mesh Network

Fig. 2.9 illustrates an example of *backbone with end-user WMNs*, where both the base stations and the end users play a role of wireless relays to forward neighboring nodes' traffic. That is, the end users are also capable of routing and self-organization.

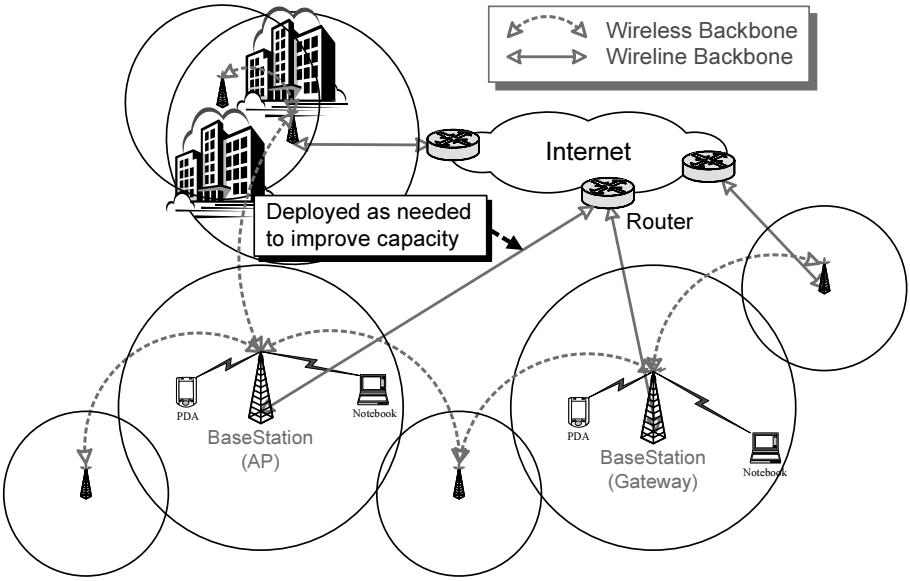



**Fig. 2.8.** Backbone wireless mesh network.

The end-user WMNs can improve the coverage of base station and network connectivity, thereby reducing the infrastructure costs due to fewer base stations needed. Noteworthily, the mobility issue in the end-user WMNs is challenging, since the network topology and connectivity will frequently change as users move. The mobility issue in end-user WMN includes seamless handoff, fast route selection, network organization and management.

### 2.2.3 Relay-Based Wireless Mesh Network

Fig. 2.10 shows an example of *relay-based wireless mesh networks*. The relay in this WMN acts as the *lightweight* BS/AP, which permits an economical design for the relays. The relaying systems can employ either *amplify-and-forward* or *decode-and-forward* schemes. In the amplify-and-forward scheme, the relays simply function as analog repeaters, thereby augmenting their own noise levels. In general, the relays in WMNs will operate in a decode-and-forward fashion. The relays can be digital repeaters, bridges, or routers, all of which will completely decode and encode the received signals before forwarding.

The objectives of deploying relays are to extend the coverage as well as to improve user throughput. If the density of relays is high enough, all the users can be served by nearby relays with a very short separation distance, thereby enhancing the link capacity between the relays and users. Then, the goals of robust and uniform data rate in the wireless networks can be achieved in a more economical way.

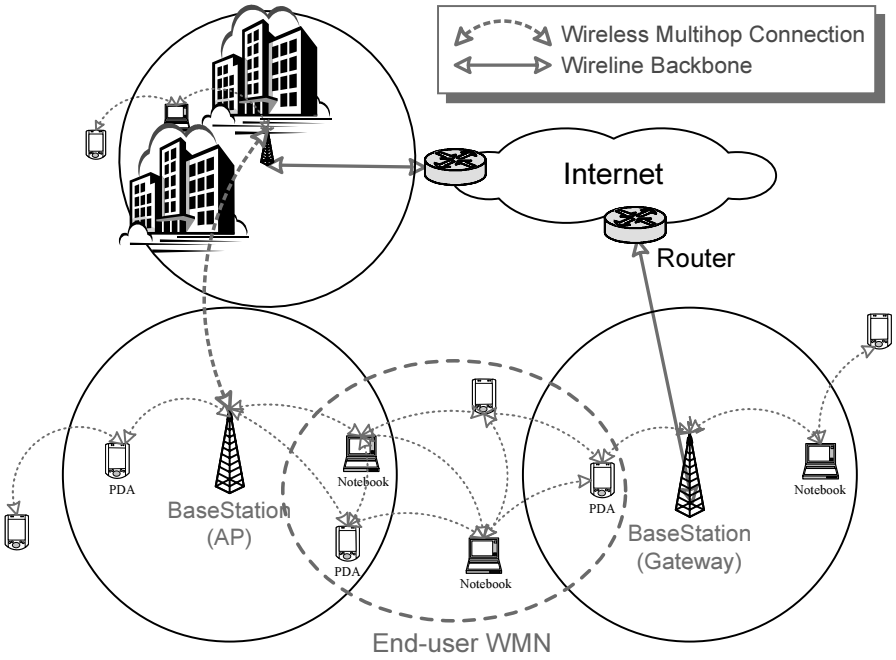



**Fig. 2.9.** Backbone with end-user wireless mesh network.

## 2.3 IEEE 802.11s Mesh Networking Technology

The IEEE 802.11 standards aim at defining the physical (PHY) layer and the MAC sublayer protocols for the wireless local area network. The IEEE 802.11b can achieve the peak rate of 11 Mbps, while the IEEE 802.11a/g WLANs achieve 54 Mbps. Furthermore, the IEEE 802.11e addresses the quality of service (QoS) issue, and the 802.11n intends to provide a data rate in excess of 200 Mbps. However, the IEEE 802.11a/b/e/g/n standards mainly focus on the one-hop infrastructure-based communications, where the stations (STAs) are directly connected to the APs. Due to lack of a scalable distributed MAC protocol, the legacy IEEE 802.11 WLANs will face the scalability issue that degrades the throughput severely in the multi-hop communications.

Therefore, the IEEE 802.11s task group (TG) is established to address the multi-hop issue for WLAN. This TGs aims to standardize the meshed WLANs by defining the PHY and MAC layer protocols to support broadcast/multicast/unicast transmissions under self-configured mesh network topology. In the IEEE 802.11s network, the WLAN mesh is defined as a set of mesh points interconnected via wireless links with the capabilities of automatic topology learning and dynamic path selection [13]. Fig. 2.11 shows an example of IEEE 802.11s WLAN mesh. In the figure, there are two classes of wireless nodes. The *mesh points* (MPs) are the nodes supporting wireless mesh services, such as mesh routing selection and forwarding, while the *non-*

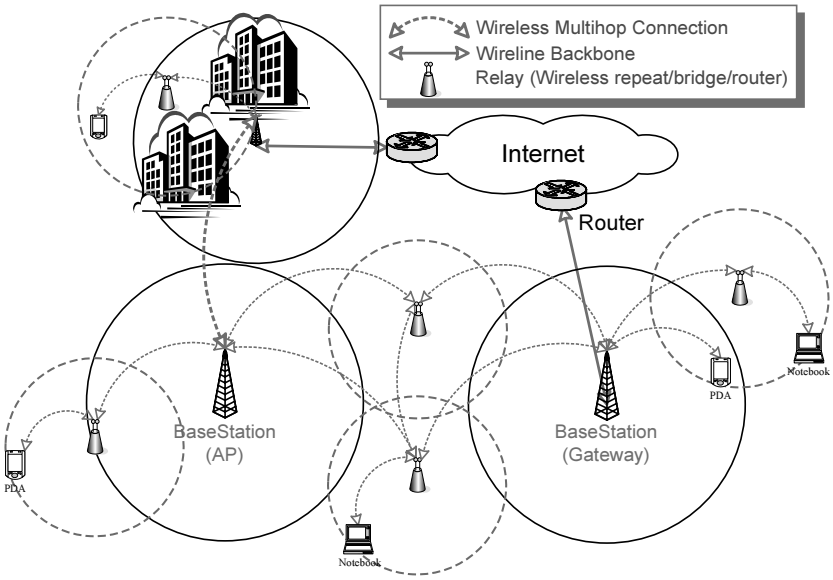



**Fig. 2.10.** Relay-based wireless mesh network.

*mesh nodes* are the pure client STAs. In addition to mesh services, the mesh access point (MAP) also provides wireless access services. The pure client STAs do not participate in the WLAN mesh, but they can associate with the mesh APs to connect to the mesh networks. The WLAN mesh can connect to other networks by the mesh portals (MPPs). Multiple WLAN meshes can also be connected by the MPP.

The IEEE 802.11s employs the IEEE 802.11e enhanced distributed channel access (EDCA) as the basis of the medium access mechanism. The enhanced MAC derived from the legacy 802.11 standard is compatible with the existing WLAN devices. To improve the network throughput and channel efficiency in the multi-hop communications, the intra-mesh congestion control and the multi-channel common channel framework (CCF) are suggested in the IEEE 802.11s [13]. By implementing a simple hop-by-hop congestion control mechanism at each MP, the intra-mesh congestion control can relieve the local congestion problem. This mechanism includes three essential elements, including the local congestion monitoring, the congestion control signaling, and the local rate control. The basic idea of the intra-mesh congestion control is to actively monitor the local channel utilization, and detect the local congestion. Through the congestion control signaling, a node can notify the upstream-hop nodes and the neighboring nodes of the local congestion. Once receiving the congestion notification, the nodes will employ the local rate control to relieve the congestion. The CCF framework provides the multi-channel MAC operation for the MP with single/multiple radio interfaces in order to boost the overall network capacity with multiple channels. In CCF, the MP in backoff will exchange the RTS/CTS-like channel negotiation message with the destination node. After suc-

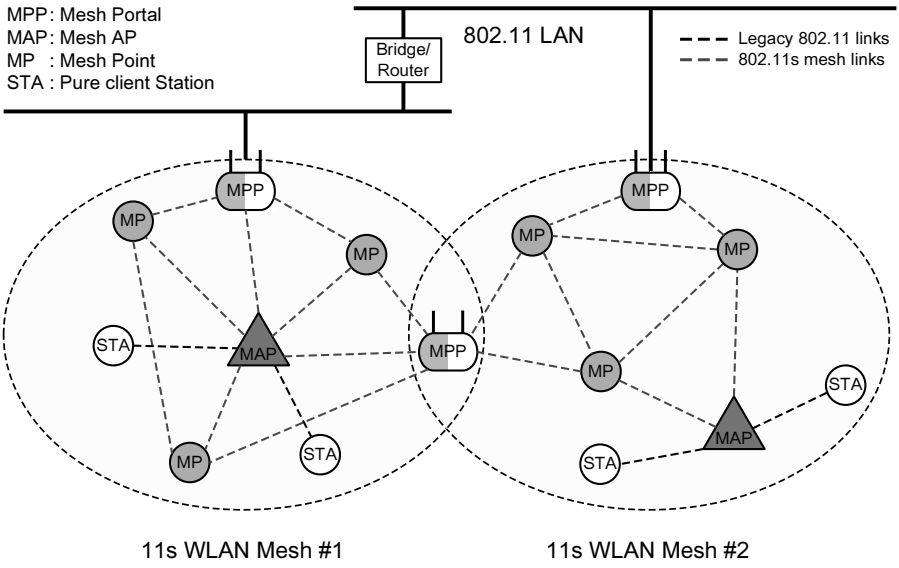



**Fig. 2.11.** The network architecture for the IEEE 802.11s WLAN mesh network.

cessful channel negotiation, MP pairs switch to the agreed channel to send/receive the data and acknowledge (ACK) frames. One advantage of CCF is that it can accommodate the legacy channel access mechanisms. That is, the common control channel for the nodes without supporting the CCF will appear as a traditional 802.11 channel.

In the IEEE 802.11s, the default hybrid wireless mesh protocol (HWMP) combines the flexibility of reactive on-demand route discovery and the efficiency of proactive routing [13, 14]. Specifically, the reactive on-demand mode in HWMP is based on the radio-metric ad hoc on-demand distance vector (RM-AODV) protocol, while the proactive mode is implemented by the tree-based routing. Such a combination in HWMP can achieve the optimal and efficient path selection. In addition, the HWMP can support various radio metrics in the path selection, such as throughput, QoS, load balancing, power-aware, etc. The default metric is the *airtime cost*, which considers the PHY and MAC protocol overhead, frame payload, and the packet error rate to reflect the radio link condition. To conclude, supporting the hybrid reactive and proactive schemes with a variety of radio metrics, the HWMP has an appealing benefit of flexibility and can be applied to a wide range of application scenarios, including fixed to mobile mesh networks.

## 2.4 IEEE 802.16 Mesh Networking Technology

The IEEE 802.16 WirelessMAN standard aims to define the PHY and MAC layer protocols to provide the broadband wireless services in the metropolitan area environment [15]. This standard supports the *point-to-multipoint* (PMP) broadband com-

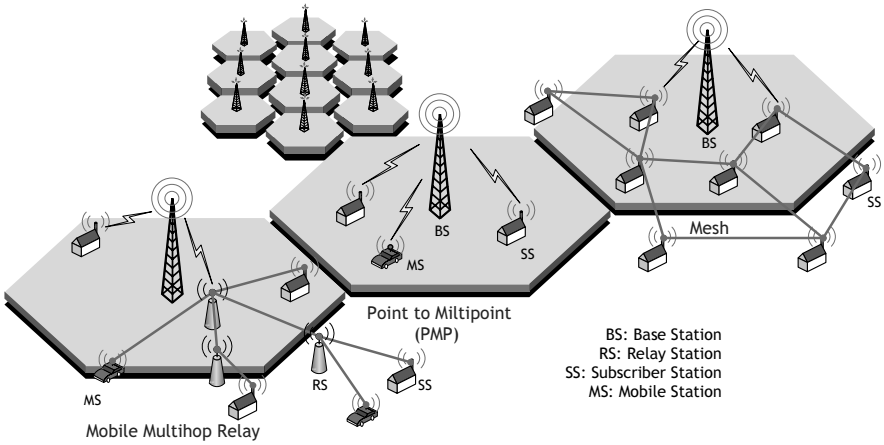



**Fig. 2.12.** An example of IEEE 802.16 networks. *Middle*: point-to-multipoint (PMP) mode. *Right*: mesh mode. *Left*: mobile multihop relay (MMR) mode.

munications, which operates in the licensed 10-66 GHz frequency band and requires the line-of-sight (LoS) link between the BS and the subscriber station (SS). In addition to the PMP mode, the IEEE 802.16a extension introduces the *mesh mode* to the IEEE 802.16 networks [16]. The mesh mode uses the lower frequency band of 2-11 GHz and allows the non-line-of-sight (NLoS) communications.

Fig. 2.12 shows an example of IEEE 802.16 network. In the figure, the SS in the PMP mode has to directly connect to the BS. On the contrary, the SS can communicate with the neighboring SSs in the mesh mode (see Fig. 2.12). Furthermore, the SS in the mesh mode can act as the wireless relay to forward others' traffic toward the central BS. Consequently, the coverage of BS can be extended, so that the infrastructure costs is substantially reduced.

However, the currently-developed mesh mode in IEEE 802.16 standard is not compatible with the original PMP mode. In the physical layer, the mesh mode has different frame structures and only supports the OFDM operation in both licensed and unlicensed bands. In the MAC layer, the network entry procedure in the mesh mode is also different. In addition, the mesh mode does not support the mobility of SS. Therefore, the IEEE 802.16 working group (WG) establishes the "*Mobile Multihop Relay (MMR)*" study group (SG), and then creates the 802.16j TG. The TG-j intends to enhance the normal PMP frame structure and develop the new relay networking protocols, with the goals of coverage extension and throughput enhancement. Different to the mesh mode, the MMR mode in the IEEE 802.16j extension focuses on efficiently providing the multi-hop relay connections between SSs/mobile stations (MSs) and the BS with a tree topology, as shown in Fig. 2.12. The MMR mode is required to be backward compatible to the PMP mode, and will support both the OFDMA and OFDM operations.

To design a practical mobile multihop relay system, many important issues still need to be addressed, including the enhanced frame structure, backward-compatible network entry procedure, synchronization and security in the multi-hop communications. To support the 802.16e MSs, the mobility management, the seamless handoff, the optimal and fast multi-hop route selection are essential issues in the 802.16j MMR systems. As for the radio resource management in MMR systems, the main challenges include interference management, spectrum efficiency, frequency reuse strategy, and scheduling policy.

## 2.5 Deployment Strategies for Scalable Wireless Mesh Networks

This section addresses the key challenge in WMN — the scalability issue from a network deployment perspective. We propose two scalable-WMN deployment strategies for the dense-urban and wide-area scenarios [11, 12].

### 2.5.1 Related Works

First, we discuss the issue of AP placement in WMNs for dense-urban coverage. Most works were based on the architecture that all the access points are directly connected to the Internet through cables [17]- [21]. In [17], an integer linear programming (ILP) optimization model was proposed for the access point placement problem, where the objective function was to maximize the signal level in the service area. In [18], an optimization approach was proposed to minimize the areas with poor signal quality and improve the average signal quality in the service area. The authors in [19] and [20] proposed optimization algorithms to minimize average bit error rate (BER). In [21], the AP deployment problem was also formulated as an ILP optimization problem with the objective of minimizing the maximum of channel utilization to achieve load balancing. In [17]- [21], the concept of wireless multi-hop communication was considered.

With respect to the performance issues for wireless mesh networks, it has been studied mainly from two directions [1]- [2], [22]- [25]. On one hand, from a coverage viewpoint, authors in [22] compared the coverage performance of a multi-hop WMN with that of a single-hop infrastructure-based network by simulations. On the other hand, from a capacity viewpoint, it was shown in [23] and [24] that the throughput per node in a uniform multi-hop ad hoc network is scaled like $O(1/\sqrt{k \log k})$, where $k$ is the total number of nodes. Moreover, the authors in [2] showed that the achievable throughput per node in a multi-hop WMN will significantly decrease as $O(1/k)$ due to the bottleneck at the central gateway. To resolve the scalability issue of multi-hop network, authors in [25] proposed a multi-channel WMN to improve the network throughput. Fewer papers considered both the capacity and coverage performance issues for a WMN, except for [1] in a single-user case. The scalability issue of WMN was not well addressed in [1]- [2], and [22]- [25].

**Table 2.1.** Link data rates versus coverage ranges for the IEEE 802.11a/b WLANs.

(a) Transmission performance of IEEE 802.11a

| Data link rate (Mbps) | 54 | 48 | 36 | 24 | 18 | 12 | 9 | 6 |
|---|---|---|---|---|---|---|---|---|
| Indoor range* (m) [26] | 13 | 15 | 19 | 26 | 33 | 39 | 45 | 50 |
| Outdoor range* (m) [26] | 30 | | | | 180 | | | 304 |
| Link capacity† (Mbps) [27] | 27.1 | 25.3 | 21.2 | 15.7 | 12.6 | 9.0 | 7.0 | 4.8 |

* 40 mW with 6 dBi gain patch antenna.

† PER = 10% and packet length = 1500 octets.

(b) Transmission performance of IEEE 802.11b.

| Data link rate (Mbps) | 11 | 5.5 | 2 | 1 |
|---|---|---|---|---|
| Indoor range§ (m) [26] | 48 | 67 | 82 | 124 |
| Outdoor range§ (m) [26] | 304 | | | 610 |

§ 100 mW with 2.2 dBi gain patch antenna.

### 2.5.2 Scalable Cluster-based Wireless Mesh Network for Dense-Urban Coverage

#### Architecture and Assumptions

This section presents the cluster-based WMN in the dense-urban area as shown in Fig. 2.5. In each cluster, only the central $AP_0$ has the wireline connection to the Internet. Other APs ae connected with wireless links. By this cluster-based WMN, the WLAN system can be rapidly deployed in the urban area with less cabling engineering work.

Specifically, in the proposed cluster-based WMN, the IEEE 802.11a WLAN standard is mainly used for data forwarding between APs, while the IEEE 802.11b/g is for data access between APs and user terminals. Recall that the IEEE 802.11a WLAN are assigned with eight non-overlapping channels for outdoor applications in the spectrum of 5.25 to 5.35 GHz and 5.725 to 5.825 GHz, whereas the IEEE 802.11b/g WLAN has three non-overlapping channels in the spectrum of 2.4 to 2.4835 GHz. To avoid the co-channel interference, frequency planing is applied to ensure two buffer cells between the two co-channel APs. Thus, the inter-cell co-channel interference is reduced and will not be considered in this work.

To deploy the WMN in a dense-urban environment, the coverage range of an AP is a key parameter. Table 2.1 shows the relationship between coverage range and link capacity for both the IEEE 802.11a/b WLANs [26]. Actually, these coverage ranges may vary depending on the environments. However, the proposed optimization approach is general enough to evaluate the performance of WMN with the various coverage ranges in different environments.

A. Throughput Model between Access Points

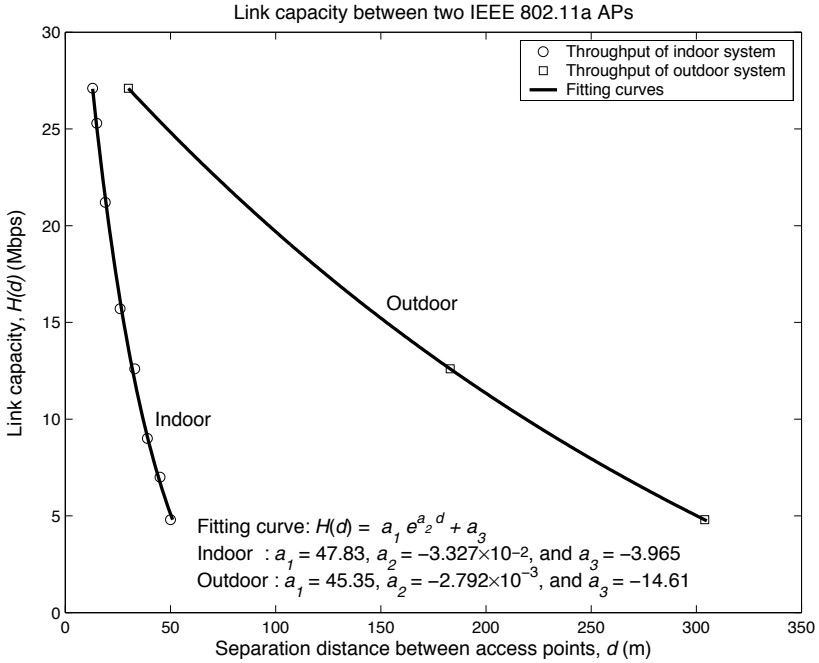



**Fig. 2.13.** The outdoor/indoor 802.11a link capacity performance $H(d)$ at a separation distance between access points $d$.

The throughput model between two APs follows the IEEE 802.11a WLAN specifications. Table 2.1 (a) lists the coverage range and link capacity for the IEEE 802.11a WLAN [26, 27]. As shown in Fig. 2.13, the radio link capacity $H(d)$ is a function of the separation distance $d$.

In this WMN, the maximum separation distance between two APs is limited by the maximum reception distance $d_{max}$. In addition, since the access points are mounted on the streetlamps, the separation distance $d$ between access points should be $d = \Omega L_S$, where $\Omega$ is a positive integer and $L_S$ is the separation distance between streetlamps.

### B. Throughput Model between an AP and Users

The design of cell size in WMN for urban coverage can be considered from two folds. First, the maximum cell radius should be less than $l_{max}$ to maintain an acceptable data rate. Second, the cell radius should be larger than $l_{min}$ to lower the handoff probability.

In each cell, users share the medium and employ the carrier sense multiple access with collision avoidance (CSMA/CA) MAC protocol to communicate with an AP. We assume that the users are uniformly distributed on the road with density $D_M$ (users/m). If the cell coverage (in radius) is $l$, the average number of users in a cell is $k = 2lD_M$. According to the method in [28], the cell saturation throughput $R_b(k)$

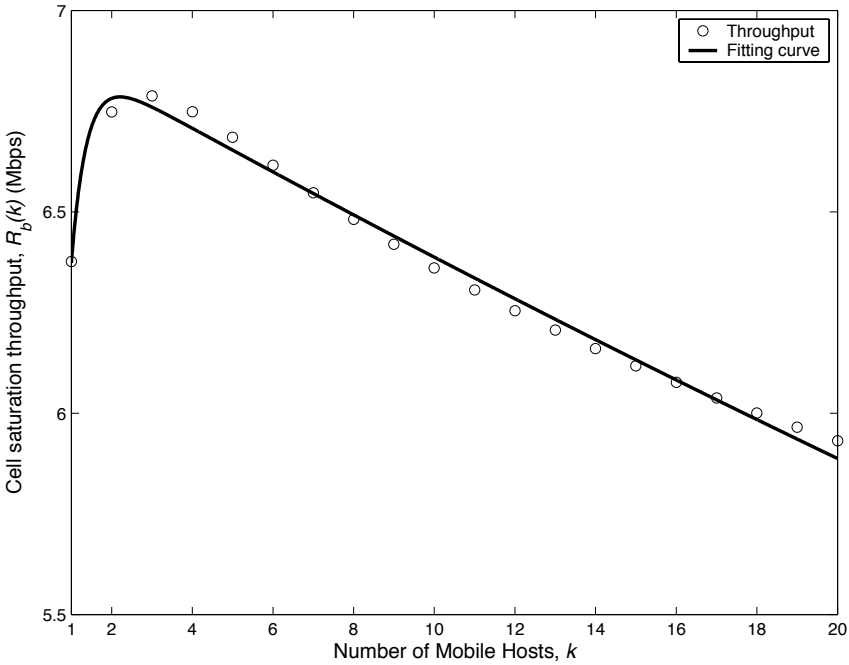



**Fig. 2.14.** The cell saturation throughput versus the number of users for the IEEE 802.11b WLAN.

of the IEEE 802.11b WLAN for various numbers of users $k$ is shown in Fig. 2.14, where data rate is 11 Mbps and average packet payload is 1500 bytes.

## Optimal Access Point Placement

### A. Problem Formulation

Radio link throughput and coverage are two essential factors in placing APs in a WMN for dense-urban coverage. From the view point of coverage, a larger cell is preferred because less number of APs are required. From the standpoint of throughput, however, a smaller cell size will be better since it can achieve a higher data rate in the wireless link. In this work, we formulate an optimization problem to determine the best separation distance for APs with consideration of these two factors.

Fig. 2.15 illustrates an example of the cluster-based WMN. Since access points will be symmetrically deployed to the central access point $AP_0$ in a cluster, only one side of the cluster needs to be considered. The notations in Fig. 2.15 are explained as follows:

– $n$ : the number of APs in the single side of the cluster;
– $d_i$ : the separation distance between $AP_{i-1}$ and $AP_i$;
– $H(d_i)$ : the radio link capacity between $AP_{i-1}$ and $AP_i$ at a distance $d_i$, according

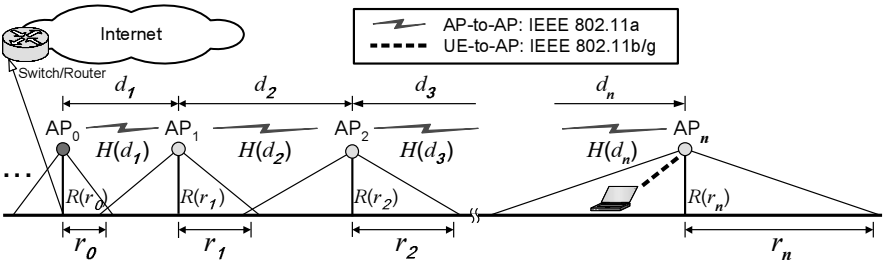



**Fig. 2.15.** A cluster of APs in the dense-urban environment (this is an example for the increasing-spacing placement strategy, where $d_1 \leq d_2 \leq \cdots \leq d_n$).

to the

IEEE 802.11a WLAN specification;
– $l_i$ : the cell radius of $AP_i$;
– $R(l_i)$ : the aggregated traffic load from all the users associated to $AP_i$, in which $R(l_i) = 2l_i D_M R_D$ and $R_D$ is the average demanded traffic of each user.
Clearly, the separation distance between two APs can be written as

$$d_i = l_i + l_{i-1}, \quad \text{for } i = 1, 2, \ldots, n \tag{2.1}$$

and the aggregated traffic load in a cell should be constrained by the cell saturation throughput, i.e.,

$$R(l_i) \leq R_b(k). \tag{2.2}$$

In the considered scenario as depicted in Fig. 2.15, the total service area in a cluster of APs is $[2l_0 + 2\sum_{i=1}^{n} 2l_i]$. Therefore, the total carried traffic load of a cluster of APs through the wireline connection can be given as

$$2\left[l_0 + 2\sum_{i=1}^{n} l_i\right] D_M R_D.$$

The total cost for deploying a cluster of APs with one wireline connection is $(2n + 1 + \rho)$, which includes the total cost of $(2n + 1)$ access points and the fixed overhead cost due to the wireline connection $\rho$. For convenience, in this work the wireline overhead $\rho$ has been normalized by the cost of one access point.

In this work, the AP placement problem will be formulated as a mixed-integer nonlinear programming (MINLP) problem with the following decision variables: $n$ and $l_0, l_1, \ldots, l_n$. The objective is to maximize the ratio of the total carried traffic load to the cost for a cluster of APs. In the following, we discuss the two AP placement strategies: the increasing-spacing and the uniform-spacing placement strategies.

### B. Increasing-Spacing Placement Strategy

Fig. 2.15 illustrates an example for the proposed increasing-spacing placement strategy, where $d_1 \leq d_2 \leq \cdots \leq d_n$. In a cluster, the aggregated carried traffic load of the wireless link between $AP_{i-1}$ and $AP_i$ is a decreasing function of $i$. That is, the further the $AP_i$ from the central $AP_0$, the less the carried traffic load in the wireless link between $AP_{i-1}$ and $AP_i$. Accordingly, it is expected to deploy access points with increasing separation distance (i.e., $d_1 \leq d_2 \leq \cdots \leq d_n$) to deliver a higher traffic load for a cluster of APs. The system parameters according to the increasing-spacing AP placement strategy can be obtained by solving the following MINLP optimization problem:

$$\underset{n,l_0,l_1,\ldots,l_n}{\mathbf{MAX}} \frac{\text{Total carried traffic load in a cluster of APs}}{\text{Total cost for deploying a cluster of APs}} = \frac{2\left[l_0 + 2\sum_{i=1}^{n} l_i\right] D_M R_D}{(2n+1+\rho)} \tag{2.3}$$

**subject to**

$$2l_i D_M R_D \leq R_b(k),\ i = 1, 2, \ldots, n \tag{2.4}$$

$$H(d_i) \geq \sum_{j=i}^{n} R(l_j) = \sum_{j=i}^{n} 2l_j D_M R_D,\ i = 1, 2, \ldots, n \tag{2.5}$$

$$d_i = l_i + l_{i-1},\ i = 1, 2, \ldots, n \tag{2.6}$$

$$l_{min} \leq l_i \leq l_{max},\ i = 0, 1, \ldots, n \tag{2.7}$$

$$d_i \leq d_{max},\ i = 1, 2, \ldots, n \tag{2.8}$$

$$d_i = \Omega_i L_S,\ i = 1, 2, \ldots, n. \tag{2.9}$$

In the following, we will explain the above constrains. Constraint (2.4) means that in each cell the total carried traffic load is constrained by the cell saturation throughput. Constraint (2.5) states the condition that the radio link capacity $H(d_i)$ between $AP_{i-1}$ and $AP_i$ should be greater than the aggregate carried traffic load from the cells served by $AP_i$, $AP_{i+1}$, ..., and $AP_n$. Constraint (2.6) is the relationship between the separation distance $d_i$ and the cell radius $l_i$. Constraint (2.7) refers to the limits of cell radius, i.e., $l_{min}$ and $l_{max}$. According to (2.8), the maximum separation distance between two access points is limited to $d_{max}$. With respect to (2.9), it is a limit on the separation distance $d_i$ due to the distance between streetlamps.

### C. Uniform-Spacing Placement Strategy

Referring to Fig. 2.5, the uniform-spacing placement strategy is to make all the cells in a cluster have the same radius, and thus the access points are uniformly deployed in the service area. Therefore, there are additional constraints for this placement, i.e., $l_i = l$ and thus $d_i = d = 2l$. Accordingly, $R(l_i) = R(l)$ and

**Table 2.2.** System parameters for numerical examples.

| Symbol | Item | Nominal value |
|---|---|---|
| $D_M$ | Road traffic density | 0.08 users/m |
| $L_S$ | Distance between two street lamps | 30 m |
| $R_D$ | Traffic demand of each user | 0.2 Mbps |
| $l_{min}$ | Min. of cell radius | 45 m |
| $l_{max}$ | Max. of cell radius | 300 m |
| $d_{max}$ | Max. distance between APs | 300 m |

$H(d_i) = H(d)$. Then, the MINLP formulation of access point placement problem can be modified as

$$\underset{n,l}{\mathbf{MAX}} \frac{(2n+1) \times 2lD_M R_D}{(2n+1+\rho)} \tag{2.10}$$

**subject to**

$$R_b(k) \geq R(l) = 2lD_M R_D \tag{2.11}$$

$$H(d) \geq nR(l) = n \times 2lD_M R_D \tag{2.12}$$

$$d = \Omega L_S. \tag{2.13}$$

**Numerical Examples of Cluster-Based WMN**

We compare the performance of the increasing-spacing placement strategy and the uniform-spacing placement strategy. The system parameters in the numerical examples are summarized in Table 2.2.

Fig. 2.16 compares the achieved profits of the objective function for the increasing-spacing and the uniform-spacing placement strategies with various wireline overheads $\rho$. Fig. 2.16 demonstrates the advantage of the increasing-spacing placement strategy over the uniform-spacing placement strategy. The achieved profit of the objective function is a concave function of the number of APs, $n$, as depicted in Fig. 2.16. Therefore, there exists an optimal solution of $n$ to maximize the profit of the objective function. For example, when the wireline overhead $\rho = 4$, $n = 3$ will achieve the best performances for both placement strategies. The corresponding cell radii for the increasing-spacing placement strategy are $(l_0, l_1, l_2, l_3) = (113.3$ m, 66.7 m, 143.3 m, 156.7 m$)$ and that for the uniform-spacing placement strategy is $l = 105$ m, respectively. Accordingly, the corresponding separation distances for the increasing-spacing placement strategy are $(d_1, d_2, d_3) = (180$ m, 210 m, 300 m$)$ and that for the uniform-spacing placement strategy is $d = 210$ m, respectively. In this case, the increasing-spacing placement strategy can achieve $15\%$ higher profit of the objective function than the uniform-spacing placement strategy. In Fig. 2.16, we can also observe that the best number of APs in a cluster can vary for different strategies. When the wireline overhead $\rho = 2$, $n = 2$ will achieve the best performance for the increasing-spacing placement strategy, and $n = 1$ for the uniform-spacing placement strategy. In this case, the achieved

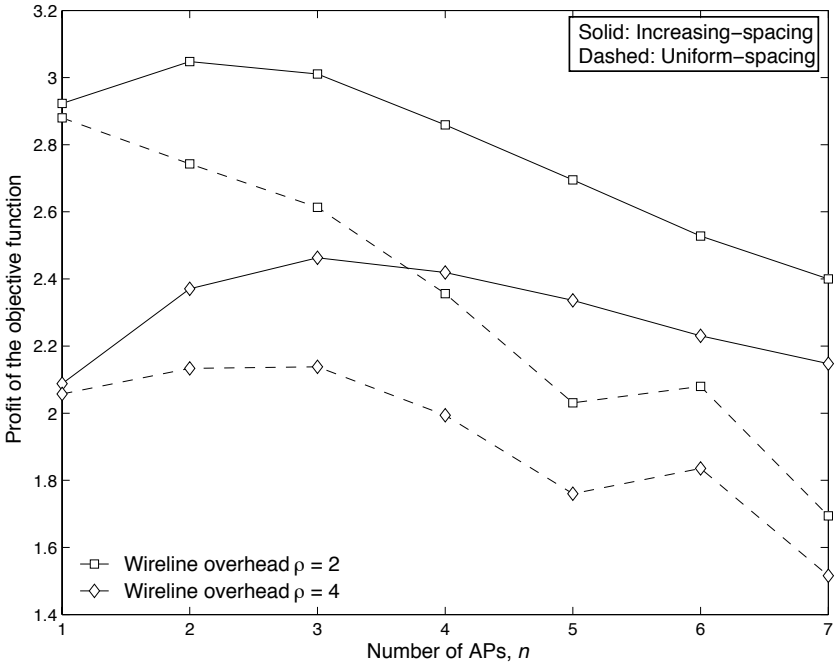



**Fig. 2.16.** Comparison of the increasing-spacing and the uniform-spacing placement strategies in terms of the achieved profit of the objective function for different wireline overheads $\rho$.

profit of the objective function for the increasing-spacing placement strategy is about $6\%$ better than that for the uniform-spacing placement strategy.

Fig. 2.17 shows the sum of carried traffic load and the total service area for a cluster of $(2n+1)$ APs according to the increasing-spacing and the uniform-spacing placement strategies. One can observe that the total carried traffic load with the increasing-spacing placement strategy increases faster than that with the uniform-spacing placement approach as the number of APs in a cluster increases. Furthermore, the increment of the traffic load for the uniform-spacing strategy will gradually diminish or even decrease (see $n = 6$ to $n = 7$). Since the profit of the objective function is proportional to the total carried traffic load, and inversely proportional to the cost of a cluster of APs, the achieved profit of the objective function is a concave function of $n$ as shown in Fig. 2.16.

### 2.5.3 Scalable Ring-Based Wireless Mesh Network for Wide-area Coverage

#### Network Architecture

Fig. 2.6 illustrates the scalable ring-based wireless mesh network for wide-area coverage. In each mesh cell, all users are connected to the central gateway in a multi-hop fashion. Each intermediate node operates as a wireless relay to forward data traffic to

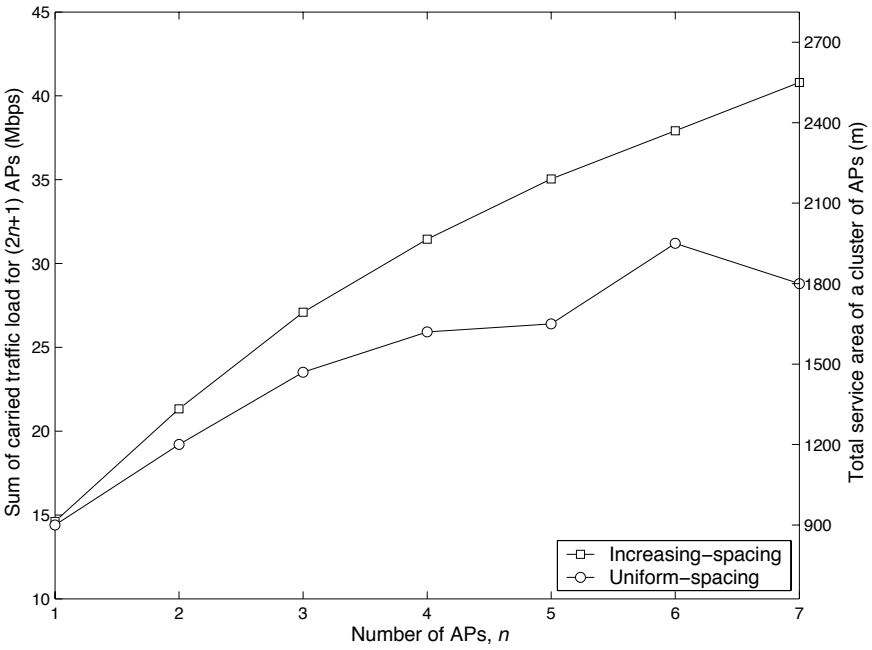



**Fig. 2.17.** Performance comparison of the increasing-spacing and the uniform-spacing placement strategies, from the viewpoint of one cluster.

the gateway. The gateway connects to the backbone network via a wired or wireless connection. Using this mesh architecture, the cabling engineering work for WMN deployment can be reduced.

In this work, we consider a multi-channel wireless mesh network. In this WMN, each mesh cell is divided into several rings, denoted by $A_i$, $i = 1, 2, \cdots, n$. The user in the ring $A_i$ will connect to the central gateway via an $i$-hop communication. We assume that each node can concurrently receive and deliver the forwarded traffic as [7, 10, 25]. That is, each node is equipped with two radio interfaces, and the users in ring $A_i$ will communicate with the users in rings $A_{i-1}$ and $A_{i+1}$ at two different channels $f_i$ and $f_{i+1}$, respectively. By doing so, the multi-hop mesh network becomes scalable to the number of users since the contention issue can be resolved by the multi-channel arrangement in a ring-based network.

We assume that frequency planning is applied to avoid the co-channel interference, and thus the inter-ring co-channel interference will not be considered in this work. In a multi-channel network [25], the dynamic frequency assignment can flexibly utilize the available channels, but it needs a multi-channel MAC protocol that is sometimes complicated. In the considered ring-based WMN, however, the fixed frequency planning is simple because it only needs to consider the width of each ring to ensure an enough co-channel reuse distance.

The carried traffic load in each mesh node includes its own traffic and the forwarded traffic from other users. Assume that all the nodes in the inner ring $A_i$ share the relayed traffic from the outer ring $A_{i+1}$. Suppose that the user density is $\rho$. The average number of nodes $c_i$ in the ring $A_i$ can be expressed as

$$c_i = \rho a_i = \begin{cases} \rho\pi r_i^2, & \text{for } i = 1 \\ \rho\pi(r_i^2 - r_{i-1}^2), & \text{for } 1 < i \leq n \end{cases} \tag{2.14}$$

where $a_i$ and $(r_i - r_{i-1})$ are the area and the width of ring $A_i$, respectively. Let $R_D$ and $R_i$ be traffic load generated by each node and the total carried traffic load per node in ring $A_i$, respectively. Then,

$$\begin{aligned} R_i &= \frac{c_{i+1}}{c_i} R_{i+1} + R_D \\ &= \left[ \frac{\sum_{j=i+1}^{n} c_j}{c_i} + 1 \right] R_D. \end{aligned} \tag{2.15}$$

For the outermost ring $A_n$, $R_n = R_D$.

**Coverage and Capacity Maximization**

A. Problem Formulation

In the following, we formulate an optimization problem to determine the best number of rings in a cell and the optimal width of each ring so as to achieve the optimal tradeoff between throughput and coverage. To begin with, we discuss the constraints in the optimization problem for the considered ring-based WMN as shown in Fig. 2.6.

- The relay link capacity $H_i(d)$ for a user in ring $A_i$ should be greater than the traffic load carried at each node $R_i$, i.e., $H_i(d) \geq R_i$, where $d$ is the separation distance between the node and the next-hop node. This constraint guarantees the minimum throughput for each user.
- The maximum reception range should be larger than the ring width $(r_i - r_{i-1})$, i.e., $(r_i - r_{i-1}) \leq d_{max} = d_1$.
- The ring width should be greater than the average distance $d_{min}$ between two neighboring nodes, i.e., $(r_i - r_{i-1}) \geq d_{min}$, where $d_{min} = 1/\sqrt{\rho}$ m is dependent on the user node density $\rho$.

B. MINLP Optimization Approach

From the above considerations, the optimal coverage issue in a wireless mesh network can be formulated as an MINLP problem with the following decision variables: $n$ (the number of rings in a mesh cell) and $r_1, r_2, \ldots, r_n$. The objective function is to maximize the coverage of a mesh cell as follows. In this scalable ring-based WMN, the ring-based frequency planning resolves the collision issue as cell coverage increases. Accordingly, the optimal coverage and capacity will be achieved simultaneously, since more users in a mesh cell can also lead to higher cell capacity. The

**Table 2.3.** System parameters for numerical examples.

| Symbol | Item | Nominal value |
|---|---|---|
| $\rho$ | User node density | $(100)^{-2}\mathrm{m}^{-2}$ |
| $R_D$ | Demanded traffic of each user node | 0.5 Mbps |
| $d_{min}$ | Min. of ring width, i.e., $(1/\sqrt{\rho})$ | 100 m |
| $d_{max}$ | Max. reception range | 300 m |
| $l_{RC}$ | Sensing range ($\gamma_I d_{max}$) | 450 m |

optimal system parameters for the ring-based WMN can be analytically determined by solving the following optimization problem:

$$\underset{n,r_1,r_2,\ldots,r_n}{\mathrm{MAX}} \; r_n \text{ (Coverage of a mesh cell)} \tag{2.16}$$

**subject to**

$$H_i(d) \geq R_i \tag{2.17}$$

$$d_{max} \geq (r_i - r_{i-1}) \geq d_{min} \tag{2.18}$$

where the cell coverage is defined as the cell radius $r_n$. A cross-layer analytical model to evaluate $H_i(d)$ was developed in [12].

### Numerical Examples of Ring-Based WMN

**Table 2.4.** Relevant network parameters for an IEEE 802.11a WLAN.

| | |
|---|---|
| PHY mode for data frame, $m_a$ | $1 \sim 8$ |
| PHY mode for control frame, $m_c$ | 1 (6Mbps) |
| Propagation Delay, $\delta$ | 1 $\mu$s |
| SIFS | 16 $\mu$s |
| DIFS | 34 $\mu$s |
| Empty slot time, $\sigma$ | 9 $\mu$s |
| $m_{bk}$ | 6 |
| Initial Contention Window, $W$ | 16 |

The system parameters are summarized in Tables 2.3 and 2.4. We considers a simple case where all the ring widths in a cell are the same, i.e., $(r_i - r_{i-1}) = r$. The control frames (RTS/CTS/ACK frames) are transmitted with PHY mode $m_c = 1$ for reliability. The mesh nodes are uniformly distributed with density $\rho = (100)^{-2}$ nodes/m. We assume the sensing range $l_{RC} = \gamma_I d_{max}$, where $\gamma_I$ is 1.5. As in [29], the chosen data frame payload sizes for eight PHY modes are {425, 653, 881, 1337,

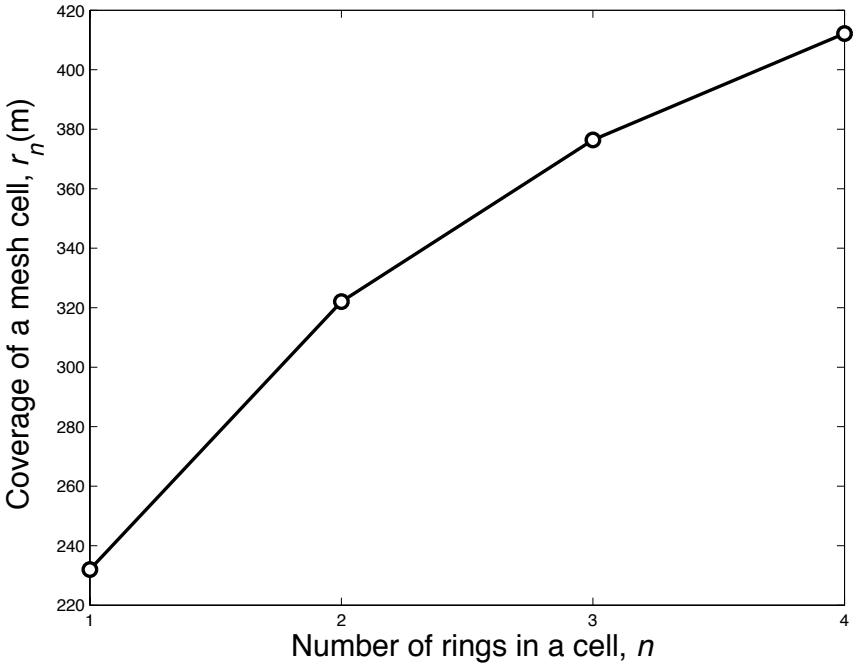



**Fig. 2.18.** Cell coverage versus the number of rings $n$ in a mesh cell, where the demanded traffic per user is $R_D = 0.5$ Mbps.

1793, 2705, 3617, 4067 $(4095 - MAC_{hdr} - MAC_{FCS})\}$ bytes. Referring to the measured results [26], the corresponding average reception ranges are $d_j = \{300$, 263, 224, 183, 146, 107, 68, $30\}$ m. It is true that these reception ranges vary for different environments. However, the proposed optimization approach is general enough to evaluate the performances of different WMNs by adopting various reception ranges.

In Fig. 2.18, the achieved cell coverage against the number of rings in a mesh cell for $R_D = 0.5$ Mbps is shown. One can observe that the optimal achieved cell coverage is 412 m with $n = 4$. Compared with the coverage of the single-hop network $(n = 1)$, the multi-hop mesh network improves the coverage by $77\%$. Fig. 2.19 illustrates the capacity performance against the number of rings in a cell, for $R_D = 0.5$ Mbps. In this example, the corresponding optimal cell throughput is 26.7 Mbps with $n = 4$. Compared with $n = 1$, the multi-hop mesh network improves the cell throughput by $215\%$.

Figs. 2.18 and 2.19 show that the proposed ring-based WMN can enhance the cell coverage and throughput compared with the single-hop network. More importantly, we find that the optimal number of rings is equal to $n = 4$ for $R_D = 0.5$ Mbps. In these figures, it is shown that the more the number of rings in a mesh cell, the better the coverage and capacity. However, the constraints on the mesh link throughput and the separation distance between the mesh nodes determine the optimal solution.

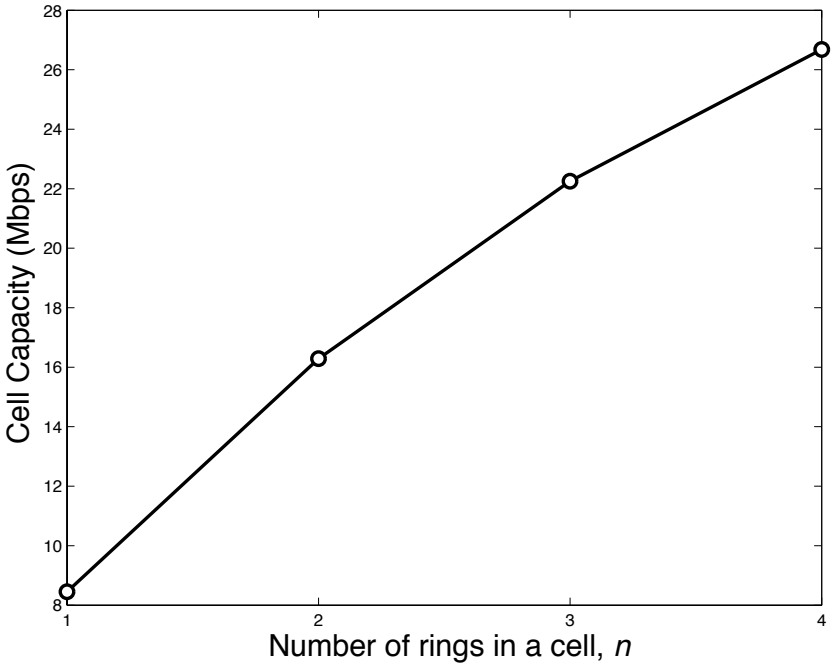



**Fig. 2.19.** Achieved cell capacity versus the number of rings $n$ in a mesh cell, where $R_D = 0.5$ Mbps.

Fig. 2.20 shows the ring width for various number of rings $n$ in a cell. Referring to this figure, when the number of rings increases, the ring width decreases. In general, when the number of rings $n$ in a cell increases, the cell coverage also increases as shown in Fig. 2.18. For handling the increment of relay traffic as $n$ increases, each ring width will decrease to shorten the hop distance and thus improve the link capacity. However, since the ring width should be larger than the average distance between two neighboring nodes, there exists a maximum value of $n$. In this example, the maximum allowable number of rings in a mesh cell is $n = 4$.

## Conclusion

Wireless mesh networking is a promising solution for the next-generation communication system to support ubiquitous broadband services with low transmission power. In this chapter, we have provided a brief overview on the mesh networking technologies for the IEEE 802.11s and IEEE 802.16 systems. Then, we address the key challenge in WMN — the scalability issue from a network deployment perspective. We present two scalable-WMN deployment strategies for the typical WMN application scenarios, including the dense-urban and wide-area scenarios. The proposed WMNs are scalable in terms of coverage, since the frequency planning with multiple

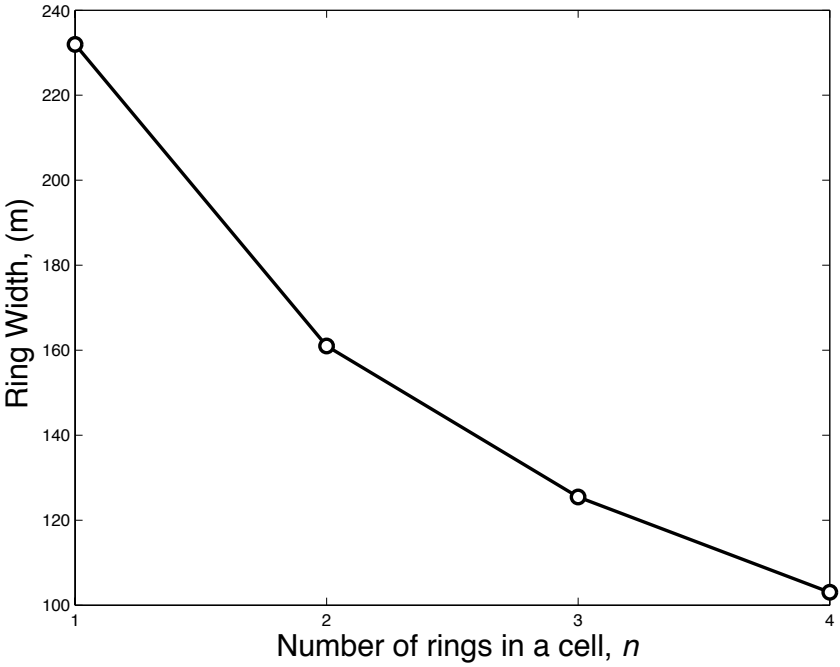



**Fig. 2.20.** Ring width $r$ versus the number of rings $n$ in a cell, where $R_D = 0.5$ Mbps.

available channels can effectively resolve the contention issue and thus the throughput can be ensured by properly designing the deployment parameters. This chapter also investigates the optimal tradeoff between capacity and coverage for the scalable WMNs. We have applied the mixed-integer nonlinear programming (MINLP) optimization approach to determine the optimal deployment parameters, subject to the tradeoffs between throughput and coverage.

## Acknowledgment

This work was supported in part by the MoE ATU Program, the Program for Promoting Academic Excellence of Universities (Phase I and II), and the National Science Council under Grant 95W803C, Grand EX-91-E-FA06-4-4, Grant NSC 95-2752-E-009-014-PAE, Grant NSC Grant NSC 95-2221-E-009-155.

## References

1. R. Pabst *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, Sept. 2004.

2. J. Jun and M. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Commun. Mag.*, vol. 10, no. 5, pp. 8–14, Oct. 2003.
3. M. J. Lee, J. Zheng, Y.-B. Ko, and D. M. Shrestha, "Emerging standards for wireless mesh technology," *IEEE Wireless Commun. Mag.*, vol. 13, no. 2, pp. 56–63, Apr. 2006.
4. M. Zhang and R. Wolff, "Crossing the digital divide: Cost-effective broadband wireless access for rural and remote areas," *IEEE Commun. Mag.*, vol. 42, no. 2, pp. 99–105, Feb. 2004.
5. T. Fowler, "Mesh networks for broadband access," *IEE Review*, vol. 47, no. 1, pp. 17–22, Jan. 2001.
6. *MeshNetworks website*, http://www.meshnetworks.com.
7. *MeshDynamics website*, http://www.meshdynamics.com.
8. B. Lewis, "Mesh networks in fixed broadband wireless access," IEEE C80216-03.10r1, July 2003.
9. L. Qiu *et al.*, "Troubleshooting multihop wireless networks," Microsoft Research Tech. Report, MSR-TR-2004-11, Nov. 2004.
10. I. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *Computer Networks*, vol. 47, pp. 445–487, Mar. 2005.
11. J.-H. Huang, L.-C. Wang, and C.-J. Chang, "Deployment strategies of access points for outdoor wireless local area networks," in *Proc. IEEE VTC'05 Spring*, May 2005.
12. ——, "Capacity and QoS for a scalable ring-based wireless mesh network," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2070–2080, Nov. 2006.
13. *Joint* SEE-M*esh/*WI-M*esh Proposal to 802.11* TG*s*, IEEE 802.11-06/328r0, Feb. 2006.
14. *Hybrid Wireless Mesh Protocol (HWMP) Overview*, IEEE 802.11-06/329r3, July 2006.
15. *IEEE Std. 802.16,* IEEE *standard for Local and Metropolitan Area Networks*, 2001.
16. *IEEE Std. 802.16-2004,* IEEE *standard for Local and Metropolitan Area Networks*, 2004.
17. R. C. Rogrigues, G. R. Mateus, and A. A. F. Loureiro, "On the design and capacity planning of a wireless local area network," in *Proc. IEEE/IFIP NOMS'00*, Apr. 2000, pp. 335–348.
18. M. Amenetsky and M. Unbehaun, "Coverage planning for outdoor wireless LAN systems," in *Proc. IEEE Zurich Seminar on Broadband Communications*, Feb. 2002.
19. M. Kobayashi *et al.*, "Optimal access point placement in simultaneous broadcast system using OFDM for indoor wireless LAN," in *Proc. IEEE PIMRC'00*, Sept. 2000, pp. 200–204.
20. T. Jiang and G. Zhu, "Uniform design simulated annealing for optimal access point placement of high data rate indoor wireless LAN using OFDM," in *Proc. IEEE PIMRC'03*, Sept. 2003, pp. 2302–2306.
21. Y. Lee, K. Kim, and Y. Choi, "Optimization of AP placement and channel assignment in wireless LANs," in *Proc. IEEE LCN'02*, Nov. 2002, pp. 831–836.
22. S. Naghian and J. Tervonen, "Semi-infrastructured mobile ad-hoc mesh networking," in *Proc. IEEE PIMRC'03*, Sept. 2003, pp. 1069–1073.
23. P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.
24. J. Li *et al.*, "Capacity of ad hoc wireless networks," in *Proc. ACM MobiCom'01*, July 2001.
25. A. Raniwala and T.-C. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. IEEE INFOCOM'05*, Mar. 2005.
26. CISCO, *Data Sheet of Cisco* A*ironet 1200 Series Access Point*.
27. J. C. Chen, "Measured performance of 5 GHz 802.11a wireless LAN systems," White Paper of Atheros Communications, Aug. 2001.

28. G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.
29. L. J. Cimini Jr. *et al.*, *Packet Shaping for Mixed Rate 802.11 Wireless Networks*, United States Patent Application Number: US 20030133427, July 2003.

# 3

# End-to-End Design Principles for Broadband Cellular Mesh Networks *

Ö. Oyman and S. Sandhu

Intel Corporation, USA
{ozgur.oyman,sumeet.sandhu}@intel.com

## 3.1 Introduction

The cellular industry is currently positioned for a transition from traditional voice networks to high-rate data networks that will enable ubiquitous internet connectivity and support of a variety of new wireless applications. The rapid deployment of broadband wireless access networks over large coverage areas (e.g., wide-area networks (WANs)) calls for the investigation of low-cost and high-performance infrastructure technologies. One major source of cost in cellular networks is the wireline backhaul that connects infrastructure devices (e.g., base stations, access points etc.) to the core internet. These wired backhaul connections are often electrical or fiber-optic and incur significant recurring costs of deployment, leasing and maintenance for service providers. Therefore, technologies that enable the wireless backhaul between the core network and infrastructure devices are of great interest from a cost savings perspective.

The demands and constraints on future wireless networks outlined above lead to a multihop *cellular mesh* architecture [1]- [2], an example of which is depicted in Fig. 3.1. The role of the additional infrastructure deployment points is to serve as *relay* terminals for the data to be *routed* between the wired infrastructure devices (labeled as BS, i.e. base station) and end users (labeled as MS, i.e. mobile station) and thereby to enhance the quality of end-to-end communication. Depending on the size of their coverage area, these fixed radio relay nodes are referred to as "micro" or "pico" relay stations (RS) (e.g., nodes 102-110 in Fig. 3.1 each cover their respective shaded hexagonal micro cells) and are generally much smaller in size and less expensive than the wired infrastructure devices. These relay deployments will serve toward various objectives, such as enhancing data rate coverage and enabling range extension over cellular networks. With this motivation, there has recently been growing interest from both academia and industry in the concept of relaying in infrastructure-

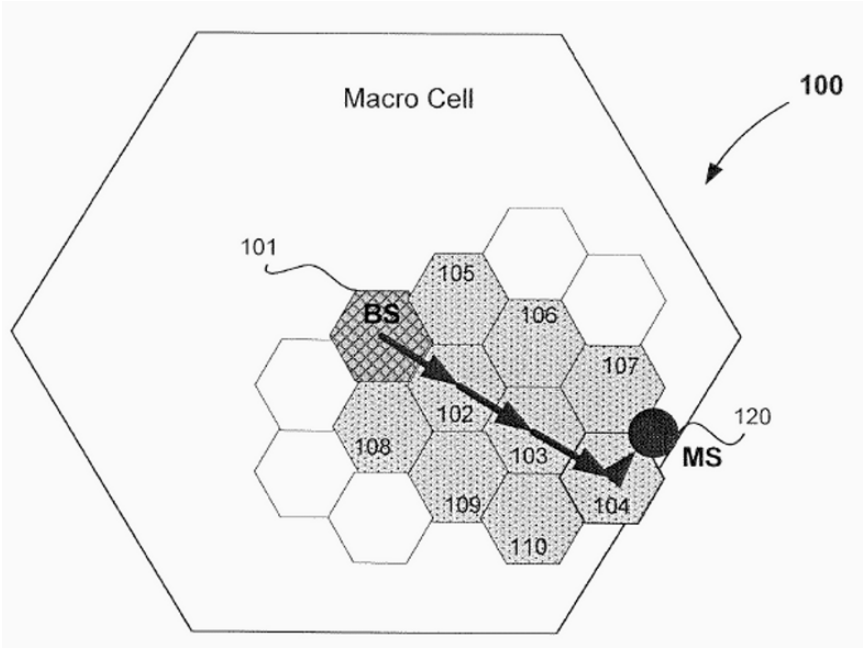



**Fig. 3.1.** Micro-cellular multihop WWAN model (© 2006 IEEE).

based wireless networks such as next generation cellular networks (B3G, 4G), wireless local area networks (WLANs) (IEEE 802.11, WiFi, HyperLAN) and broadband fixed wireless networks (IEEE 802.16, WiMax, HyperMAN).

**Chapter Overview.** End-to-end optimization of certain quality of service (QoS) measures such as throughput, reliability and latency plays a key role in designing novel algorithms and architectures for next generation relay-assisted broadband cellular mesh networks. Toward this end, the development of performance characterization methodologies as a function of the physical channel conditions and system parameters is essential to manage end-to-end QoS requirements. Motivated by the observation that both base stations and relay stations are stationary (fixed network topology) and are expected to enjoy slow-varying channel conditions which allows for rate-adaptive relaying over each hop, we first perform an information-theoretic capacity analysis to *propose end-to-end throughput and latency measures for multihop routing* as a function of the physical channel and system parameters in broadband cellular mesh networks employing orthogonal frequency division multiplexing (OFDM). Several centralized functionalities coordinated by the base station (e.g. scheduling algorithms, routing algorithms, network entry and handoff, latency management, other MAC and higher layer functions) can benefit from the knowledge of such end-to-end quality of service (QoS) measures; examples of which are provided in later sections. We have previously reported our research results in earlier publications [2]- [7] and standard contributions [8]- [9].

This chapter is organized as follows:

- **Section 3.2** introduces a broadband fading physical channel model [3] to study multihop communication protocols over cellular mesh networks and discusses our assumptions regarding channel statistics and terminal transmission/reception capabilities.
- **Section 3.3** characterizes the end-to-end capacity and throughput in a broadband cellular mesh network in the presence of a multihop routing protocol that employs rate-adaptive relaying and OFDM-based codeword transmissions over each hop [3, 5], accounting for transmission errors due to decoding failures and retransmissions until successful message reception based on an automatic repeat request (ARQ) mechanism.
- **Section 3.4** shows that end-to-end throughput maximization in a broadband cellular mesh network is equivalent to a *minimum-cost routing problem* [3] by defining the multihop routing metric as the reciprocal of the per-hop throughput, which is also known as the expected transmission time (ETT) [10], and the objective of route selection is to dynamically minimize end-to-end latency.
- **Section 3.5** discusses the use of end-to-end metrics toward the design of novel resource allocation, scheduling and multihop routing algorithms. In particular, we propose *orthogonal frequency division multihop multiple access* ($\mathrm{OFDM}^2\mathrm{A}$) resource allocation policy [4] for relay-assisted broadband cellular mesh networks, which ensures interference-free multi-user communication. This policy allows for the design of low-complexity centralized opportunistic scheduling algorithms by *separating the problems of subcarrier allocation and multihop route selection*.
- **Section 3.6** illustrates another use of end-to-end metrics for network entry and handoff; under the assumption that relay stations have capabilities very similar to base stations, i.e. they can perform association, authentication, time/frequency resource allocation. In particular, we propose novel algorithms for network entry and handoff in cellular mesh networks that yield enhanced link performance while maintaining the backward compatibility of the end users.
- **Section 3.7** briefly summarizes ongoing standardization activities in IEEE 802.16j Relay Task Group.

## 3.2 Multihop Broadband Channel Model

Consider an $N$-hop routing path between a BS and an end station (an RS or an MS), where we index each hop by $n$, such that $n = 1, ..., N$. The source terminal (e.g., for downlink, the source terminal is the BS) is identified as $\mathcal{T}_1$, the destination terminal (e.g., for downlink, the destination terminal is the end station) is identified as $\mathcal{T}_{N+1}$ and the intermediate terminals (i.e., RSs) are identified as $\mathcal{T}_2$-$\mathcal{T}_N$. Using OFDM modulation turns the frequency-selective fading channel into a set of parallel frequency-flat fading channels, rendering multi-channel equalization particularly simple since for each OFDM tone a narrowband receiver can be employed. We assume that the length of the cyclic prefix (CP) in the OFDM system is greater than

the length of the discrete-time baseband channel impulse response. This assumption guarantees that the frequency-selective fading channel indeed decouples into a set of parallel frequency-flat fading channels. We define the channel frequency response over the $n$-th hop as

$$H_n\left(e^{j2\pi\theta}\right) = \sum_{l=0}^{L-1} \sqrt{\left(\frac{\xi_n}{d_n^p}\right)} h_{n,l} e^{-j2\pi l\theta} \quad 0 \le \theta < 1$$

where we assume that the discrete-time channel has order $L-1$, $h_{n,l} \in \mathbb{C}$ represents the $l$-th tap ($l = 0, ..., L-1$) of the frequency-selective fading channel impulse response realization at hop $n$ (one can think of each of the taps representing a scatterer cluster with each of the paths emanating from within the same scatterer cluster experiencing the same delay), $\xi_n$ represents the shadow fading realization over hop $n$, $d_n$ is the inter-terminal distance between terminals $\mathcal{T}_n$ and $\mathcal{T}_{n+1}$ and $p$ is the path loss exponent. Consequently, the discrete-time complex baseband input-output relation for the frequency-flat channel over the $k$-th tone ($k = 1, ..., K$) and $n$-th hop ($n = 1, ..., N$) is given by

$$y_{n+1,k} = H_n\left(e^{j2\pi(k/K)}\right) s_{n,k} + z_{n+1,k}$$

where $K$ is the total number of OFDM tones (subcarriers), $s_{n,k} \in \mathbb{C}$ is the scalar data input signal transmitted from terminal $\mathcal{T}_n$ over the $k$-th tone and $n$-th hop, satisfying the average transmit power constraint $\mathbb{E}\left[|s_{n,k}|^2\right] = P$, $y_{n+1,k} \in \mathbb{C}$ is the reconstructed scalar data output signal at terminal $\mathcal{T}_{n+1}$ over the $k$-th tone and $n$-th hop, and $z_{n+1,k} \in \mathbb{C}$ is the temporally white zero-mean circularly symmetric complex additive white Gaussian noise (AWGN) signal at $\mathcal{T}_{n+1}$, independent across $n$ and $k$, satisfying $\mathbb{E}\left[z_{n,k} z^*_{n',k'}\right] = \sigma_z^2 \delta[k-k'] \delta[n-n']$, where $\delta[n] = 1$ if $n = 0$ and $\delta[n] = 0$, $n \neq 0$ and $\sigma_z^2$ is the variance of the discrete-time noise process $z_{n,k}$. It should be noted that due to the presence of multipath delay spread, the individual channel gains $H_n\left(e^{j2\pi(k/K)}\right)$ for $k = 1, ..., K$ will be correlated.

We write each of the taps as the sum of a fixed (possibly line-of-sight (LOS)) component, $\overline{h}_{n,l} = \mathbb{E}[h_{n,l}]$ and a variable (or scattered) component $\tilde{h}_{n,l}$ as $h_{n,l} = \overline{h}_{n,l} + \tilde{h}_{n,l}$. The channel over the $n$-th hop is said to be Rayleigh fading if $\overline{h}_{n,l} = 0$ for $l = 0, ..., L-1$ and Ricean fading if $\overline{h}_{n,l} \neq 0$ for at least one $l \in \{0, ..., L-1\}$. The variable components $\{\tilde{h}_{n,l}\}$ are assumed to be circularly symmetric complex Gaussian random variables with zero mean and variance $\sigma_l^2$. Different scatterer clusters are assumed to be uncorrelated (independence across $l$) and channels are assumed to be independent across multiple hops (across $n$), i.e., $\mathbb{E}[\tilde{h}_{n,l} \tilde{h}^*_{n',l'}] = 0$ if $l \neq l'$ or $n \neq n'$. We associate a Ricean K-factor with each of the taps over each hop by defining $\kappa_{n,l} = |\overline{h}_{n,l}|^2 / \sigma_l^2$. The relative strengths of the channel taps with powers $\{|\overline{h}_{n,l}|^2 + \sigma_l^2\}_{l=0}^{L-1}$ are determined by a certain power delay profile (PDP), which is fixed across all hops over a given routing path (i.e., across $n$).

Regarding message transmissions over the multihop link, the fading states are assumed to remain constant during the transmission of a codeword and that the channel

coherence time is much larger than the coding blocklength (slow fading assumption). Due to slow fading, each terminal in the multihop network is considered to obtain full channel state information (CSI) regarding its neighboring links (transmit CSI available through a feedback link), i.e., terminal $\mathcal{T}_n$ has perfect CSI for the links from itself to terminals $\mathcal{T}_{n-1}$ and $\mathcal{T}_{n+1}$. At each hop, the perfect CSI knowledge (or even partial CSI knowledge through the feedback of the instantaneous signal-to-noise ratios (SNRs)) at the transmitters allows for instantaneous rate adaptation to changing channel conditions on a codeword by codeword basis.

## 3.3 Information-Theoretic Characterization of Capacity in OFDM-Based Multihop Networks

We shall take an information-theoretic approach toward the analysis of the end-to-end throughput performance and deal with the capacity behavior of OFDM-based relay-assisted multihop networks in broadband slow-fading environments. As terminals can often not transmit and receive at the same time in the same frequency band, we only focus on time-division based (half duplex) relaying for the capacity analysis. In particular, we consider a simple $N$-hop decode-and-forward routing protocol, where, at hop $n$, relay terminal $\mathcal{T}_{n+1}$, $n = 1, ..., N-1$ hears and fully decodes the entire codeword transmitted from terminal $\mathcal{T}_n$ and forwards its re-encoded version to terminal $\mathcal{T}_{n+2}$. Under the described multihop routing protocol and assumptions on the channel characteristics stated in Section 3.2, the end-to-end (instantaneous) conditional mutual information $I$ (as a function of the random fading channel parameters) of the broadband cellular mesh network can be expressed as a function of the transmit signal-to-noise ratio $\mathsf{SNR}$ in the form

$$I(\mathsf{SNR}) = \max_{\sum_{n=1}^{N} \lambda_n = 1} \min \{\lambda_n I_n(\mathsf{SNR})\} \tag{3.1}$$

where $\lambda_n \in [0, 1]$ is the fractional time of the channel corresponding to hop $n$ and $I_n(\mathsf{SNR})$ is the conditional mutual information (as a function of path loss, shadowing and fading) over hop $n$ [2]. This mutual information is achievable by the multihop network under *optimal time-sharing and rate adaptation* to instantaneous shadowing and (slow) fading variations. Evaluating (3.1), we obtain the following harmonic mean formula [5, 6]:

$$\begin{aligned} I(\mathsf{SNR}) &= \frac{\prod_{n=1}^{N} I_n(\mathsf{SNR})}{\sum_{m=1}^{N} \prod_{m \neq n} I_m(\mathsf{SNR})} \\ &= \left( \sum_{n=1}^{N} \frac{1}{I_n(\mathsf{SNR})} \right)^{-1}. \end{aligned} \tag{3.2}$$

[2] The reader is referred to [12, 13] for further capacity results on non-fading AWGN multihop networks.

For the broadband frequency-selective channel over each hop, the use of OFDM modulation allows for the application of a narrow-band receiver for each tone $k = 1, ..., K$ and hence each subchannel can be viewed as a frequency-flat fading link, which implies that $I_n(\mathsf{SNR})$ can be written as [14]

$$I_n(\mathsf{SNR}) = \frac{1}{K}\sum_{k=1}^{K} I_{n,k}(\mathsf{SNR}) \tag{3.3}$$

where $I_{n,k}$ is the mutual information for the frequency-flat channel over the $k$-th OFDM tone and $n$-th hop (ignoring the loss in spectral efficiency due to the presence of the CP), which, under the assumption of Gaussian inputs, i.e., $s_{n,k} \in \mathbb{C}$ has the temporally i.i.d. zero-mean circularly symmetric complex Gaussian distribution, is given by (in bits per second per Hertz (bps/Hz))

$$I_{n,k}(\mathsf{SNR}) = \log_2\left(1 + \mathsf{SNR}\left|H_n\left(e^{j2\pi(k/K)}\right)\right|^2\right) \tag{3.4}$$

such that $\mathsf{SNR} = \frac{P}{K\sigma_z^2}$, and hence total transmit power $P$ over the $n$-th hop is shared equally across all $K$ OFDM tones. In the presence of practical system limitations due to finite number of supportable code rates and modulation sizes, the mutual information in (3.4) may not be achievable and a loss in mutual information may be incurred. In this setting, the mutual information over each hop can be computed assuming finite modulation sizes (no Gaussian inputs) and further discretized considering a finite number of code rates that guarantees a certain level of reliability (e.g., packet error rate (PER)), i.e. when using practical link adaptation mechanisms which are designed to optimize performance under such reliability and delay constraints.

To implement this rate-adaptive multihop relaying solution over random time-varying channels (e.g., fading wireless channels), where each element in the conditional mutual information set $\{I_n\}_{n=1}^N$ can be treated as a random variable, the transmit terminal over hop $n$ only needs to know the value of $I_n$ and the value of an end-to-end link quality parameter $M$, which is defined as $M = \sum_{n=1}^{N} 1/I_n$. The knowledge of global CSI (i.e. CSI for all links in the multihop network) is not required at every terminal [7], which implies significantly reduced messaging overhead. The information on $I_n$ can be obtained by each terminal through CSI feedback from only the neighboring terminal. Due to the stationarity of the infrastructure devices, the channels experienced over all hops are expected to be slowly time-varying (except possibly for the last hop involving the end user) and therefore it is realistic to assume that each node will be able to track its transmit/receive channels and perform rate-adaptive relaying. On the other hand, the parameter $M$ depends on the channel conditions over all links, which may be computed in a distributed fashion using a routing algorithm (e.g., destination-sequenced distance-vector (DSDV) [11]) in which the cost of the link over hop $n$ is represented by the metric $1/I_n$, which is also known as the *expected transmission time (ETT)* [10] in the networking literature. Such a distributed approach involves the end-to-end propagation of a single parameter, only requiring neighbor-to-neighbor message passing of the accumulated

multihop link cost metric which is updated by each terminal with the addition of the cost of the last hop. Once the total route cost $\sum_{n=1}^{N} 1/I_n$ has been determined by one of the end terminals, the value of $M$ can be broadcasted to all the terminals in the linear multihop network. Again, due to slow fading, it can be safely assumed that the update broadcasts of this parameter do not need to be performed frequently, ensuring low complexity in the protocol overhead.

Using the capacity-based performance measures for the multihop routing protocols summarized so far, we shall now investigate merits of relay-assisted mesh techniques through several numerical examples under practically relevant cellular communication settings. For purposes of illustration, unless stated otherwise, we will assume that the base station, relay stations and mobile station are equi-distantly aligned to form a linear multihop network.

**Example 1: Cell Edge Spectral Efficiency Enhancement.** We study the communication scenario depicted in Fig. 3.2, where a cell edge user could be supported by either (i) direct transmission/reception ($N = 1$) by the base station (BS-MS link), or (ii) relay-assisted communication over two hops ($N = 2$) through BS-RS and RS-MS links. Table 3.1 lists our assumptions for the channel and system parameters for a 7-cell network in the downlink mode (1 center cell and 6 interfering neighbor cells). In particular, we consider a high-capacity line-of-sight (LOS) BS-RS link, non-LOS channels for BS-MS and RS-MS links based on the Erceg-Greenstein (EG) path loss model [15] and lognormal shadowing. We also consider frequency-selective fading under the broadband channel model of Section 3.2; where each multipath fading link has four independent taps ($L = 4$) with an exponential power delay profile (PDP) and complex Gaussian (Ricean) distribution with mean $1/\sqrt{2}$ and variance $1/2$, i.e., $\kappa_{n,l} = 1$, $\forall n$, $l = 0, 1, 2, 3$. While we keep the RS-MS distance fixed at $0.3$ km, we allow the user's distance from the BS (and hence BS-RS distance) vary in order to characterize spectral efficiency as a function of range, where, consistently with Table 3.1, we take $\mathrm{Range(BS \rightarrow MS)} = \mathrm{Range(BS \rightarrow RS)} + 0.3$ (in km). Table 3.2 presents the spectral efficiency comparison between direct communication with mutual information $I_{\mathrm{direct}}$ and relay-assisted two-hop routing with end-to-end mutual information $I_{\mathrm{relay}} = 1/(1/I_1 + 1/I_2)$, where quantities $I_1$ and $I_2$ represent the mutual information over the BS-RS and RS-MS links, respectively (following (3.2)-(3.4)). We observe from these results that under favorable LOS conditions over the BS-RS link, multihop relaying boosts capacity and coverage gains against direct transmission for cell edge users suffering from poor signal-to-interference-plus-noise ratio (SINR) conditions. For example, we find in Table 3.2 that a multiplicative spectral efficiency gain factor of $4$ can be achieved for a 0 dB user at range $2.2$ km with two-hop relaying.

**Example 2: Reliability Enhancement against Fading.** In this example, we focus on the benefits of multihop for mitigating fading, which may at first seem counter-intuitive. For the purposes of the following numerical study, we will consider the broadband channel model of Section 3.2 with frequency-selective multipath fading and path-loss, but without shadowing and channel distributions are assumed to be uniform across all hops (which was not the case in Example 1). Each multipath fading link has two independent taps ($L = 2$) with an exponential power de-

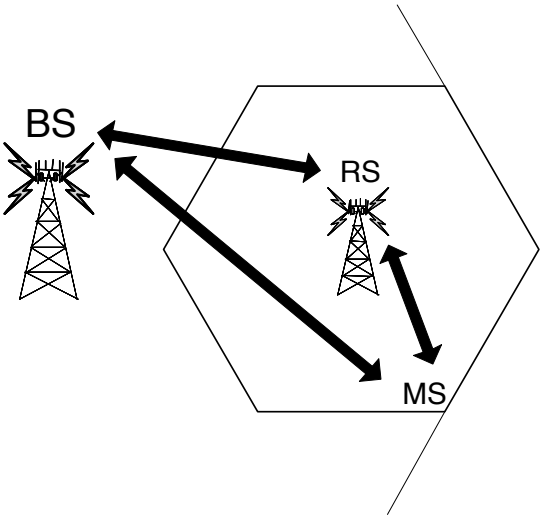



**Fig. 3.2.** Cell edge users could be either (i) directly served by BS (BS-MS link) or (ii) served through a relay-assisted two-hop route including BS-RS and RS-MS links.

lay profile (PDP) and complex Gaussian (Ricean) distribution with mean $1/\sqrt{2}$ and variance $1/2$, i.e., $\kappa_{n,l} = 1$, $\forall n$, $l = 0, 1$. The path loss exponent is assumed to be $p = 4$, and the average received SNR between the mobile user and the base station is normalized to $0\,\mathrm{dB}$. We plot in Fig. 3.3 the cumulative distribution function (c.d.f.) of the end-to-end mutual information for both fixed-rate and rate-adaptive multihop relaying schemes [6] with varying number of hops $N = 1, 2, 10$. We observe that with increasing number of hops, the c.d.f. of mutual information sharpens around the mean (i.e. the probability distribution function (p.d.f.) *concentrates*), yielding significant enhancements at low outage probabilities [16] over single-hop communi-

**Table 3.1.** Cellular mesh link budget for BS-MS, BS-RS and RS-MS channels.

| Parameter | BS-MS Channel | BS-RS Channel | RS-MS Channel |
|---|---|---|---|
| Transmit power (RMS) (dBm) | 36 | 36 | 30 |
| Transmitter gain (dBi) | 6 | 6 | 6 |
| Receiver gain (dBi) | 0 | 6 | 0 |
| Noise PSD (dBm/Hz) | -167 | -167 | -167 |
| Bandwidth (MHz) | 20 | 20 | 20 |
| Path loss model | EG | LOS | EG |
| Shadowing std (dB) | 8 | 4 | 8 |
| Tx antenna height (m) | 25 | 25 | 12 |
| Rx antenna height (m) | 2 | 12 | 2 |
| Carrier frequency (GHz) | 3.5 | 3.5 | 3.5 |
| Cell radius (km) | 1-3 | 1-3 | 0.3 |

**Table 3.2.** Spectral efficiency comparison of direct transmission vs. two-hop relaying for cell edge users.

| Range(BS → MS) (km) | $I_{\text{direct}}$ (bps/Hz) | $I_{\text{relay}}$ (bps/Hz) |
|---|---|---|
| 1 | 0.9 | 1.1 |
| 1.6 | 0.4 | 1.0 |
| 2.2 | 0.2 | 0.8 |

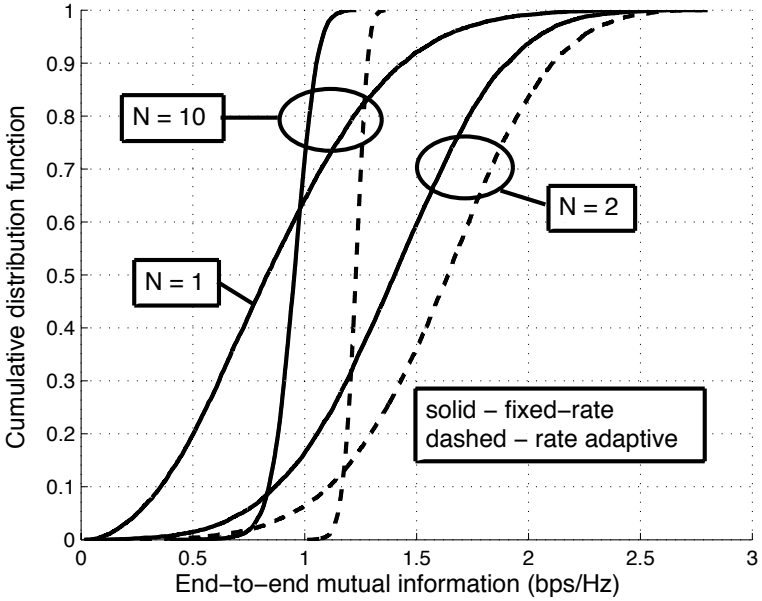



**Fig. 3.3.** Cumulative distribution function of end-to-end mutual information for fixed-rate and rate-adaptive multihop relaying schemes for $N = 1, 2, 10$ (© 2006 IEEE).

cation. We interpret this improvement of the link robustness as *multihop diversity*, which serves to ensure higher reliability in diversity-limited fading environments as well as for QoS-constrained and delay-limited applications. It was shown in [6] that, for any given desired level of end-to-end data rate $R$, there exists an optimal number of hops that minimizes end-to-end outage probability and this optimal number increases with decreasing $R$. Furthermore, [7] investigated the performance advantages from multihop relaying under an end-to-end delay constraint and identifies the conditions under which a better rate-reliability-delay tradeoff can be achieved over singlehop communication.

**Example 3: Sensitivity to Physical Channel Characteristics.** In this example, we consider the impact of varying path loss, which depends closely on range,

antenna heights, terrain characteristics and carrier frequency, on the end-to-end mutual information of an OFDM-based broadband cellular mesh network specified by (3.2)-(3.4). As the path loss characteristics of the network change with respect to the choice of these system design parameters, the optimal number of hops to maximize end-to-end mutual information would also vary, and consequently an important question is the sensitivity of the optimal route-length on the design parameters. Considering realistic broadband wireless channel models [15], preliminary simulation results (see [8] for further results) are sufficient to show the high sensitivity of gains from multihop routing to various channel parameters. In Fig. 3.4, we analyze the expected value of the optimal number of hops, denoted as $N_{opt}$, as a function of the path loss exponent $p$ assuming rate-adaptive relaying along with an end-to-end average received SNR of $0$ dB between the base station and end station, lognormal shadowing of standard deviation values $\sigma_s = 0, 4, 8$ dB (uniform across all hops) and a frequency-selective channel model with 2 independent exponential PDP taps and complex Gaussian (Ricean) fading distribution with mean $1/\sqrt{2}$ and variance $1/2$, i.e., $\kappa_{n,l} = 1$, $\forall n, l$. We average the optimal number of hops over various fading realizations using Monte Carlo simulations. Clearly, these results show the high sensitivity of $N_{opt}$ with changing $p$ and $\sigma_s$, necessitating the use of accurate channel models in order to extract the highest gains from multihop cellular mesh system designs. The only regime in which $N_{opt}$ appears to be robust with respect to $p$ and $\sigma_s$ is for high path loss exponent range, e.g., $p > 4$.

## 3.4 End-to-End Throughput and Latency over ARQ-Supported Multihop Network

Using the information theoretic characterization of end-to-end capacity in Section 3.3, we can provide further insights toward the end-to-end performance of multihop routing over broadband cellular mesh networks in terms of throughput and latency, by introducing an ARQ mechanism applicable over any given hop upon decoding failures due to transmission errors. In this setting, an ARQ protocol is considered, where, upon detection of codeword error (e.g., practical systems typically use a cyclic redundancy check (CRC) code), the erroneous codeword is discarded by the receiver and the retransmission of the codeword is requested from the transmitter. At any given hop, the retransmission request is repeated until the decoder detects an error-free transmission. It is assumed that the channel states $\{\{h_{n,l}\}_{n=1}^{N}\}_{l=0}^{L-1}$ do not change during retransmissions.

The communication in the broadband cellular mesh network under the described ARQ-supported multihop routing protocol can be characterized using a finite-state Markov chain model [7] with a discrete-time stochastic process $Z_j$, $j = 1, 2, ...$, which has $N + 1$ states indexed by $n = 1, ..., N + 1$ as depicted in Fig. 3.5, where $j$ is the code block transmission index. The transition from state $n$ to state $n + 1$ represents the transmissions from terminal $\mathcal{T}_n$ to terminal $\mathcal{T}_{n+1}$ over hop $n$ and each transmission could result in a success which means that the Markov chain arrives at state $n + 1$ or in a failure which means that the Markov chain remains at state $n$. The

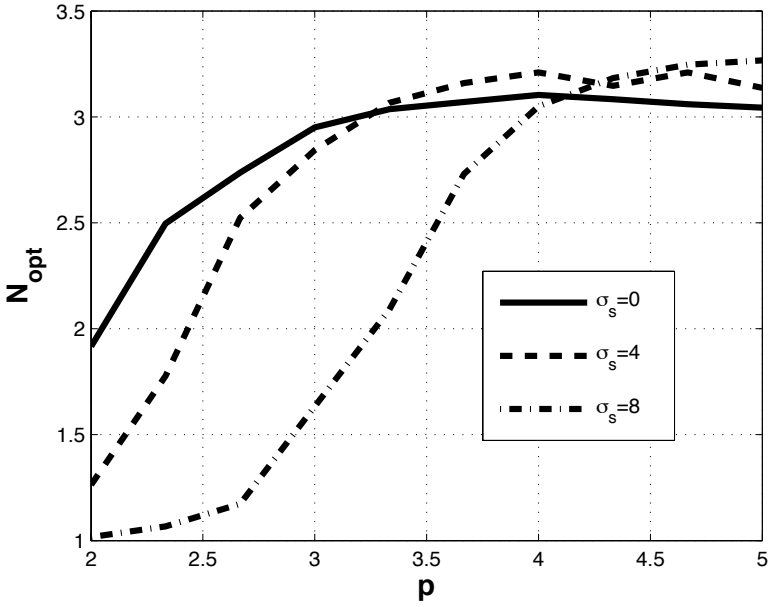



**Fig. 3.4.** The expected value of the optimal number of hops $N_{opt}$ as a function of the path loss exponent $p$ for different values of shadowing standard deviation, $\sigma_s = 0, 4, 8$ dB (© 2006 IEEE).

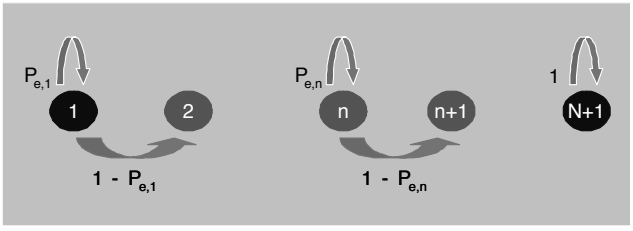



**Fig. 3.5.** Markov chain model to characterize communication under ARQ-supported multihop routing in a broadband cellular mesh network (© 2006 IEEE).

state-transition probabilities are functions of the codeword error probabilities $P_{e,n}$, which stay constant over retransmissions (since $\{\{h_{n,l}\}_{n=1}^{N}\}_{l=0}^{L-1}$ do not change). Arrival at state $N+1$ implies successful decoding of the message by the destination terminal $\mathcal{T}_{N+1}$, and thus this state is modeled as an absorption state; which means that the Markov chain terminates upon entering this state (i.e., no more transmissions are necessary), whereas states $1, ..., N$ are transient.

Our objective is to use the described Markov chain model to compute the expected value of end-to-end latency $T$ (in seconds) until the successful reception of a codeword carrying $B$ bits of information by the destination terminal assuming OFDM-based rate-adaptive broadband transmissions. In this setting, the set of data rates $\{R_n\}_{n=1}^N$ (in bits/second) is determined over hops $n = 1, ..., N$ on the basis of the knowledge of a fixed set of channel realizations $\{\{h_{n,l}\}_{n=1}^N\}_{l=0}^{L-1}$ and target codeword error probabilities $\{P_{e,n}\}_{n=1}^N$, such that $R_n$ is the highest data rate that can be supported by the channel over hop $n$ while the target codeword error probability $P_{e,n}$ is satisfied. Toward this goal, we define the stopping time $J$ of the Markov process as

$$J = \min\{j \geq 1 : Z_j = N + 1 \,|\, Z_1 = 1\}$$

based on which $T$ can be represented as

$$T = \mathbb{E}\left[\sum_{j=1}^{J-1} \frac{B}{R_{Z_j}} \mid Z_1 = 1\right].$$

This expectation can easily be computed by using the well-known first-step analysis technique based on the application of the law of total probability, exploiting the Markov property of the process $Z_j$. Now, defining $T_n$ to be the expected number of channel usage until the message arrives at state $N + 1$ given that the message is currently at state $n$, expressed as

$$T_n = \mathbb{E}\left[\sum_{j=1}^{J-1} \frac{B}{R_{Z_j}} \mid Z_1 = n\right], \; n = 1, ..., N + 1$$

we can specify the end-to-end expected latency over the multihop network for the set of data rates $\{R_n\}_{n=1}^N$ by the set of recursive relations (by conditioning on the outcome of the next transmission)

$$\begin{aligned} T_n &= (T_n + \frac{B}{R_n})P_{e,n} + (T_{n+1} + \frac{B}{R_n})(1 - P_{e,n}) \\ &= \frac{B}{R_n} + T_n P_{e,n} + T_{n+1}(1 - P_{e,n}), \quad n = 1, ..., N \end{aligned}$$

under the constraint $T_{N+1} = 0$. Solving for $T$, we obtain

$$T = \sum_{n=1}^{N} \frac{B}{R_n(1 - P_{e,n})}$$

as the expected value of end-to-end latency of multihop communication for the set of per-hop data rates $\{R_n\}_{n=1}^N$ chosen to meet the target codeword error probabilities $\{P_{e,n}\}_{n=1}^N$ for the set of channel states $\{\{h_{n,l}\}_{n=1}^N\}_{l=0}^{L-1}$. Consequently, the expected value of end-to-end throughput $R$ in bits/second is given by the harmonic mean formula [3, 9]

$$R = \frac{B}{T} = \left( \sum_{n=1}^{N} \frac{1}{R_n(1 - P_{e,n})} \right)^{-1}$$

which is consistent with (3.2). Moreover, the analysis provides beneficial insights for designing routing metrics for dynamic end-to-end QoS management over broadband multihop cellular mesh networks. Defining the cost of each link as the reciprocal of the per-hop throughput, we obtain the routing metric $\pi_n$ given by

$$\pi_n = \frac{1}{R_n(1 - P_{e,n})}. \tag{3.5}$$

The throughput-maximizing (or latency-minimizing) routing path is the path that minimizes total cost given by $\sum_{n=1}^{N} \pi_n$ (it is clear that $N$ will vary from path to path) [3, 9]. We note that the routing metric given in (3.5), known as ETT, has been proposed earlier in the context of mesh-based wireless local area networks (WLANs) in [10] and that we consider it for end-to-end QoS tracking and optimization in cellular WWANs. Any routing algorithm (e.g., destination-sequenced distance-vector (DSDV) algorithm [11]) may be executed to find the path that maximizes $R$ (or minimizes $T$).

In the presence of a more advanced ARQ mechanism (e.g. chase combining) for which the codeword error probabilities improve upon retransmissions, the finite state Markov chain model for the ARQ-supported multihop routing protocol uses state-transition probabilities that are functions of $\{\{P_{e,n,t}\}_{n=1}^{N}\}_{t=1}^{\infty}$, where $P_{e,n,t}$ denotes the conditional codeword error probability over hop $n$ during transmission $t = 1, 2, ...$ (given that transmissions over the first $t - 1$ trials were unsuccessful) such that $P_{e,n,u} < P_{e,n,v}$ for all $u > v$ and these probabilities remain constant for any given codeword transmission since the channel states $\{\{h_{n,l}\}_{n=1}^{N}\}_{l=0}^{L-1}$ do not change during retransmissions. Performing a similar analysis as before based on the first-step analysis technique, we find that the expected value of the end-to-end latency can be expressed as

$$T = B \sum_{n=1}^{N} \frac{1 + \sum_{m=1}^{\infty} \prod_{l=1}^{m} P_{e,n,l}}{R_n}$$

and the corresponding expected end-to-end throughput would be given by the formula [3, 9]

$$R = \left( \sum_{n=1}^{N} \frac{1 + \sum_{m=1}^{\infty} \prod_{l=1}^{m} P_{e,n,l}}{R_n} \right)^{-1}.$$

In this setting, the routing metric to optimize end-to-end QoS in terms of throughput and latency would be

$$\pi_n = \frac{1 + \sum_{m=1}^{\infty} \prod_{l=1}^{m} P_{e,n,l}}{R_n}. \tag{3.6}$$

## 3.5 Scheduling, Routing and Resource Allocation Based on End-to-End Metrics

Current and evolving standards for broadband wireless systems are adopting orthogonal-frequency division multiple access (OFDMA) as the resource allocation policy, in which the available time and frequency resources over each wireless link are orthogonally allocated across users, avoiding inter-user interference and impairments due to multipath fading. For fixed portable applications, where radio channels are slowly varying, an intrinsic advantage of OFDMA over other multiple access methods is its capability to exploit *multiuser diversity* [17]- [19] embedded in diverse frequency-selective channels while simultaneously taking advantage of channel variations over time. While OFDMA resource allocation over traditional point-to-point cellular systems is well understood [20]- [22], it is not yet clear how to extend multiuser diversity concepts achieved by opportunistic scheduling mechanisms to multihop/mesh architectures. Novel OFDMA resource allocation schemes and scheduling algorithms are essential toward the design of the multihop wireless backhaul for relay-assisted cellular systems.

Under the assumption that all users share the same bandwidth, and the channel state information for the fading channels over all multihop links and over all subcarriers is collected apriori at the BS, the problem of user scheduling and route selection can be solved jointly in a centralized fashion in order to maximize the total network capacity. In this regard, past work in [23] has taken a linear programming approach for determining the optimal routing and scheduling of flows that maximizes throughput in the context of code division multiple-access (CDMA) based multihop networks. However, such centralized joint scheduling and routing approaches may impose significant computational complexity at the BS as well as requiring fast and reliable feedback and feed forward channels for exchanging information among BS, RSs and MSs, which results in a significant overhead problem with increasing density of users and infrastructure relay terminals in network.

In this section, we consider resource allocation over an OFDMA-based cellular mesh network and design low-complexity suboptimal algorithms in which user scheduling and multihop route selection mechanisms are separated. Motivated by the observation that both BS and RSs are stationary (fixed network topology) and are expected to enjoy slow-varying physical channel conditions which simultaneously allows utilizing from opportunistic scheduling, multihop routing and rate-adaptive relaying mechanisms, we *extend OFDMA-based resource allocation to multihop communication* settings and introduce an approach based on centralized scheduling with the objective of simultaneously achieving throughput maximization, coverage extension and user fairness across the cellular mesh network with minimal computational complexity and messaging overhead.

Over the multihop downlink/uplink communication links between the BS and MSs, we assume that only a single terminal (BS, RS or MS) transmits to another terminal (BS, RS or MS) in a given cellular time/frequency resource. In other words, we extend OFDMA resource allocation to cellular mesh networks ensuring that no two simultaneous transmissions can occur over any given time slot and frequency tone

among all multihop routing paths serving the MSs and thereby totally avoiding intracell interference. The terminology $\mathrm{OFDM}^2\mathrm{A}$ refers to such orthogonal allocation of time/frequency resources across users in a multihop cellular mesh network [4].

We remark that $\mathrm{OFDM}^2\mathrm{A}$ resource allocation enables simultaneously achieving throughput and reliability gains from multi-user diversity by the opportunistic scheduling of multiple users and multi-route diversity by dynamic multihop route adaptation to changing channel conditions. Furthermore, it should be noted that $\mathrm{OFDM}^2\mathrm{A}$ ensures no intracell interference among the users in the BS coverage area through the orthogonal allocation of the available time/frequency resources. Moreover, the use of $\mathrm{OFDM}^2\mathrm{A}$ requires no significant modification to the existing point-to-point cellular architectures; for instance, the same frequency reuse patterns can be employed. All the appealing qualities of OFDMA, such as easy decoding at the processing power-limited user side at downlink, carry on to cellular mesh networks through the usage of $\mathrm{OFDM}^2\mathrm{A}$.

The resource allocation can be performed in a centralized fashion where the BS decides which end user to transmit information over multiple hops at each time slot and frequency subcarrier, or it can be performed such that the RSs can locally perform some level of resource allocation in a distributed/hybrid fashion. Our forthcoming discussion will consider a centralized resource allocation policy, where the base station is the sole decision-maker for allocating the time and frequency resources across users and the actions of the relay terminals are fully coordinated by the base station.

We now build on the $\mathrm{OFDM}^2\mathrm{A}$-based resource allocation policy by introducing a *centralized scheduling framework* at the BS in a relay-assisted cellular mesh network. As depicted in Fig. 3.6 and described over the next two paragraphs, this framework relies on the key *principle of separating subcarrier allocation and multihop route selection mechanisms* [4]; which reduces algorithm design complexity significantly resulting in simple scheduling and routing algorithms and minimal messaging overhead.

*Multihop route selection (Fig. 3.6, Step 1):* By using a distributed routing algorithm (e.g. DSDV), the capacity-optimal routes for each user can be constructed in a distributed fashion requiring a small amount of overhead. While our scheduling algorithms are also applicable in conjunction with centralized routing, distributed implementation of routing may be more favorable. This is due to two reasons; (i) distributed routing has lower overhead complexity compared with centralized routing as it only requires passing end-to-end (or aggregate) route metrics rather than per-link metrics, (ii) slow-fading physical channel characteristics allow for lower-frequency route updates and ensure the reliability of end-to-end link quality estimates. The per-link cost metrics capture the physical channel fading conditions and are chosen with the objective of choosing multihop routing paths that maximize the end-to-end capacity, see [3] for further analysis.

*Subcarrier allocation (Fig. 3.6, Step 2):* After the execution of the routing algorithm and determination of the optimal multihop routes for each user (i.e., network tree rooted at BS), the BS is informed of the lowest-cost (i.e., capacity-maximizing) end-to-end route metrics of all users over all subcarriers and uses these route metrics

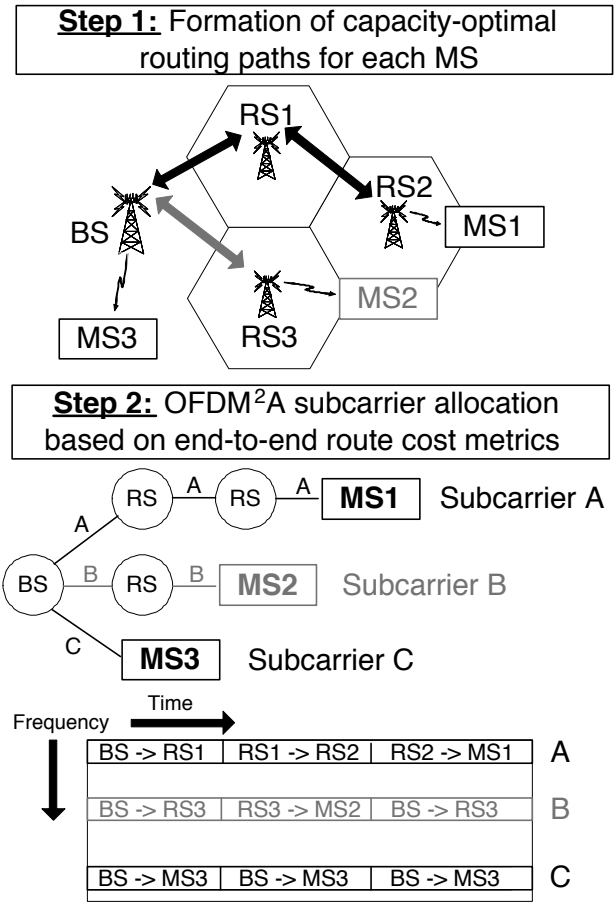



**Fig. 3.6.** $\mathrm{OFDM}^2\mathrm{A}$-based resource allocation under the novel principle of separating subcarrier allocation and multihop route selection (© 2006 IEEE).

for opportunistic scheduling by *assigning frequencies to users based on their route qualities*. Once a subcarrier is allocated to a user, this subcarrier is used over all hops in the routing path to transmit the data of the chosen user. Potential interference across multiple hops in the routing path is avoided by the orthogonal time-sharing policy enforced by $\mathrm{OFDM}^2\mathrm{A}$. It should be emphasized that the centralized scheduling algorithms only require the knowledge of the end-to-end route cost metrics at the BS, and not the per-hop cost metrics corresponding to the individual links. From these end-to-end route cost metrics, the BS can infer the instantaneous end-to-end capacity of all the users over different frequencies (as a function of the physical channel conditions over multiple hops corresponding to each user's routing path), which makes well known scheduling algorithms like max-SINR and proportional-fair scheduling applicable for subcarrier allocation over the cellular mesh network. It

was shown in [4] that these centralized scheduling algorithms simultaneously realize gains from both multiuser diversity and multihop relaying to enhance capacity and coverage, provided the availability of closed-loop transmission mechanisms.

## 3.6 Network Entry and Handoff Based on End-to-End Metrics

This section presents another usage of end-to-end metrics in the design of QoS-optimizing algorithms for multihop cellular mesh networks. Upon waking up or moving from one cell to another, the MS must decide whether to associate with a base station or a relay station, which requires determination of policies for network entry and handoff. Our approach to this problem will rely on our end-to-end capacity analysis in Section 3.3. Moreover, we will assume that relay stations have capabilities very similar to base stations, i.e. they can perform association, authentication, time/frequency resource allocation etc. with some control from the BS via the wireless backhaul links. Such relay stations look like base stations to legacy end users and can provide fully backward-compatible functionalities.

Consider the communication scenario depicted in Fig. 3.2 where the MS wishes to communicate with the BS at the highest possible throughput; where the comparison between $I_{\mathrm{direct}}$ and $I_{\mathrm{relay}}$ as defined in Example 1 of Section 3.3 will determine the optimal route: (i) If $I_{\mathrm{direct}} \geq I_{\mathrm{relay}}$, then the MS connects to the BS directly and the BS-MS link is used to convey data. (ii) If $I_{\mathrm{direct}} < I_{\mathrm{relay}}$, then the MS connects to the RS and the two-hop route composed of BS-RS and RS-MS links is used to send information. This method ensures that the MS performs network entry in a throughput-optimal fashion, while ensuring backward compatibility. The network entry is realized based on this condition through power control at the RS terminal. When RS advertises itself as a potential receiver to the MS, it lowers down its transmit power by a fraction of

$$\alpha = \left(2^{I_{\mathrm{relay}}} - 1\right)\left(2^{I_2} - 1\right)^{-1}$$

which ensures that the cost of conveying the packet over the BS-RS link is taken into account (recall from Example 1 of Section 3.3 that $I_{\mathrm{relay}} = 1/(1/I_1 + 1/I_2)$, where quantities $I_1$ and $I_2$ represent the mutual information over the BS-RS and RS-MS links, respectively). In this setting, the RS must possess the knowledge of the channel qualities over both BS-RS and RS-MS links, so that it can compute $I_1$ and $I_2$ and reduce its power by the factor $\alpha$. Finally, it should be noted that if the quality of the BS-RS link is much better than the RS-MS link (i.e. $I_1 \gg I_2$), then $\alpha \approx 1$ and accounting for the BS-RS link quality in network entry and handoff is not required.

## 3.7 Multihop Relaying in Cellular Standards

Although multihop and mesh-based wireless networking techniques have been standardized in the context of local and personal area networks (e.g., IEEE 802.11s,

IEEE 802.15.4), standardization efforts toward future cellular wide area networks have only recently begun. The multihop relay (MR) study group was formed in July 2005 to evalute merits of multihop relaying technologies for future IEEE 802.16-based wide area networks. The project authorization request (PAR) was approved in the March 2006 IEEE Standards meeting to initiate the 802.16j Relay Task Group; the standard is expected to be completed and approved in early 2008. The first phase of 802.16j is expected to be restricted to infrastructure relay stations that extend coverage of 802.16e base stations without impacting the subscriber station specification. These relay stations will be fully backward-compatible in the sense that they will operate seamlessly with existing 802.16e subscribers. Key technical topics currently discussed in the 802.16j task group include general relay concepts, frame structures, network entry, bandwidth request, handover, construction and transmission of medium access control (MAC) protocol data units (PDUs), measurement and reporting, scheduling, routing, interference control and mobility management.

## Conclusion

End-to-end optimization of certain quality of service (QoS) measures such as throughput, reliability and latency plays a key role in designing novel algorithms and architectures for next generation relay-assisted broadband cellular mesh networks. In this chapter, we presented results on the end-to-end capacity, throughput and latency characterization of multihop communication as a function of the physical channel and system parameters in broadband cellular mesh networks with special focus on transmissions under orthogonal frequency-division multiplexing (OFDM) modulation. Furthermore, we discussed the role of these end-to-end metrics in the system-level optimization of cellular mesh networks, and their use in the design of novel multihop routing, rate-adaptive relaying, scheduling and resource allocation, network entry and handoff algorithms.

## References

1. R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. Fettweis, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, Sept. 2004.
2. Ö. Oyman, J. N. Laneman, and S. Sandhu, "Multihop relaying for broadband wireless mesh networks: From theory to practice," *IEEE Communications Magazine*, 2007.
3. Ö. Oyman, "End-to-end throughput and latency measures for multihop routing in relay-assisted broadband cellular OFDM systems," in *Proc. IEEE Radio and Wireless Symposium*, Long Beach, CA, Jan. 2007.
4. Ö. Oyman, "$\mathrm{OFDM}^2\mathrm{A}$: A centralized resource allocation policy for cellular multihop networks," in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers*, Monterey, CA, Oct. 2006.

5. Ö. Oyman and S. Sandhu, "A Shannon-theoretic perspective on fading multihop networks," in *Proc. Conference on Information Sciences and Systems (CISS'06)*, Princeton, NJ, March 2006.
6. Ö. Oyman and S. Sandhu, "Non-ergodic power-bandwidth tradeoff in linear multihop networks," in *Proc. IEEE International Symposium on Information Theory (ISIT'06)*, Seattle, WA, July 2006.
7. Ö. Oyman, "Reliability bounds for delay-constrained multihop networks," in *Proc. Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep. 2006.
8. Ö. Oyman and S. Sandhu, "Throughput improvements in micro-cellular multihop networks," in *IEEE 802.16 Multihop Relay Study Group*, Vancouver, Canada, Nov. 2005, available online at: http://wirelessman.org/relay/index.html.
9. Ö. Oyman, S. Sandhu, and N. Himayat, "End-to-end throughput metrics for QoS management in 802.16j MR Systems," in *IEEE 802.16 Multihop Relay Task Group*, Dallas, TX, Nov. 2006, available online at: http://wirelessman.org/relay/index.html.
10. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio multihop wireless mesh networks," in *Proc. ACM MobiCom'04*, Philadelphia, PA, Sep. 2004.
11. C. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proc. ACM SIGCOMM*, London, UK, 1994.
12. M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. E. Fuja, "On the optimum number of hops in linear ad hoc networks," in *Proc. IEEE Inform. Theory Workshop (ITW)*, San Antonio, TX, Oct. 2004.
13. M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. E. Fuja, "Bandwidth and power efficient routing in linear wireless networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2624–2633, June 2006.
14. H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 225–234, Feb. 2002.
15. *Channel Models for Fixed Wireless Applications*, IEEE 802.16.3c-01/29r5, 2003.
16. L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, May 1994.
17. R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE Int. Computer Conf.*, Seattle, WA, Jun. 1995.
18. D. Tse and S. Hanly, "Multi-access fading channels: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.
19. P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.
20. Y. W. Cheong, R. S. Cheng, K. B. Latief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
21. D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.
22. M. Ergen, S. Coleri, and P. Varaiya, "QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems," *IEEE Transactions on Broadcasting*, vol. 49, no. 4, pp. 362–370, Dec. 2003.

23. H. Viswanathan and S. Mukherjee, "Throughput-range tradeoff of wireless mesh backhaul networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 593–602, Mar. 2006.

# 4

# Medium Access Control and Routing Protocols for Wireless Mesh Networks

J. C. Hou, K.-J. Park, T.-S. Kim, and L.-C. Kung

University of Illinois, Urbana, USA
{jhou,kjp,tskim,kung}@cs.uiuc.edu

## 4.1 Introduction

Wireless mesh networks (WMNs), a.k.a. *community wireless networks*, have emerged to be a new cost-effective and performance-adaptive network paradigm for the next-generation wireless Internet. Targeting primarily for solving the well-known *last mile problem* for broadband access [1, 2], WMNs aim to offer *high-speed coverage at a significantly lower deployment and maintenance cost*. As shown in Fig. 4.1, most of the nodes are stationary in WMNs. Only a fraction of nodes have direct access, and will serve as gateways, to the Internet. In addition, several nodes serve as relays forwarding traffic from other nodes (as well as their own traffic) and maintain network-wide Internet connectivity, while the remaining nodes send frames along dynamically selected ad-hoc paths to the gateway nodes with Internet access. WMNs are preferable to existing cable/DSL based networks or wireless LANs (that provide WiFi access), due to the following advantages: (i) mesh networks are more cost-effective as service providers do not have to install a wired connection to each subscriber ($20–$50K per square mile to establish access, approximately 1/4 of the cost incurred in high speed cable access); (ii) mesh networks are inherently more reliable since each node has redundant paths to reach the Internet; (iii) the throughput attained by a mesh network user can be, in principle, increased through routing via multiple, bandwidth-abundant paths (in contrast, in WLANs the shared bandwidth decreases as the number of users within a HotSpot increases); and (iv) WMNs can readily extend their coverage by installing additional ad-hoc hops.

Several cities are planning or have partially deployed WMNs, such as *Bay Area Wireless Users Group (BAWUG)* [3], *Champaign-Urbana Community Wireless Network* (*CUWiN*) [4], *SFLan* [5], *Seattle Wireless* [6], *Southampton Open Wireless Network (SOWN)* [7], and *Wireless Leiden* (in Netherlands) [8]. The academic/research efforts are, on the other hand, represented by the MIT *Roofnet* project [9], the Rice University *Technology for All* project [10], and the MSR *Self-organizing neighborhood wireless mesh networks* project [1]. Although initial successes have been reported in these efforts, a number of performance related problems have also been identified. Excessive packet losses [11–13], unpredictable channel behaviors [11,12],

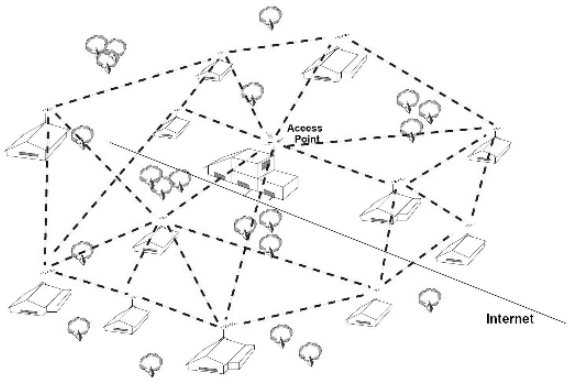



**Fig. 4.1.** Wireless mesh network.

inability to find stable and high-throughput paths [11, 12], and throughput degradation due to intra-flow and inter-flow interference [13–15] are among those most cited.

All these problems are rooted in the fact that *the notion of a link is no longer well-defined in wireless environments*. In network theory and practice, a link is usually characterized by its bandwidth, latency, packet loss ratio and patterns. However, in a WMN, a wireless medium is *shared* among nodes, and the *sharing range* is determined by (i) several PHY/MAC attributes such as the transmit power, the carrier sense threshold, and the channel on which an interface sends/receives frames, (ii) intra- and inter-flow interference (which in turn is contingent upon how nodes and traffic are distributed in the spatial and temporal domains), and (iii) environmental factors, such as multi-path fading and shadowing effects, temperature and humidity variation, and existence of objects in between. As a result, all the definitive metrics that characterize a link are no longer *well-defined* for a wireless link. All the protocols that were devised, and well-suited, for wireline networks will likely yield poor performance or even fail in WMNs. For example, as shown in [11] and [16], the shortest path algorithm with the hop count as the link metric will likely identify paths that are composed of long, lossy links with low bandwidth.

To solve (or at least mitigate) the above problems, one should (a) characterize how, and to what extent, wireless links are affected by PHY/MAC attributes and other environmental factors, (b) identify control knobs in the PHY/MAC layers with which the sharing range of a wireless link can be better controlled, and (c) understand the implication of making available these PHY/MAC attributes to higher-layer protocols on system performance optimization. Central to issues (a) and (b) is medium access control (MAC), while issue (c) is usually termed as cross layer design and optimization.

In this chapter, we discuss the state of the art in designing and implementing MAC for WMNs in Section 4.2. In particular, we categorize existing MAC-related re-

search into four categories: (1) controlling the sharing range of the wireless medium and increasing spatial reuse; (2) exploiting temporal/spatial diversity; (3) exploiting availability of multiple channels; and (4) exercising rate control. We also introduce in Section 4.3 a modular programming environment, termed as the *Transport Device Driver (TDD)*, that exports the PHY/MAC attributes via well-defined APIs and facilitates cross layer design and optimization, as a case study of cross layer design and optimization. We then present various routing protocols that take advantage of PHY/MAC attributes (such as channels) for route optimization in Section 4.4. We discuss open research issues in Section 4.5. Finally, our conclusion follows.

## 4.2 Medium Access Control in WMNs

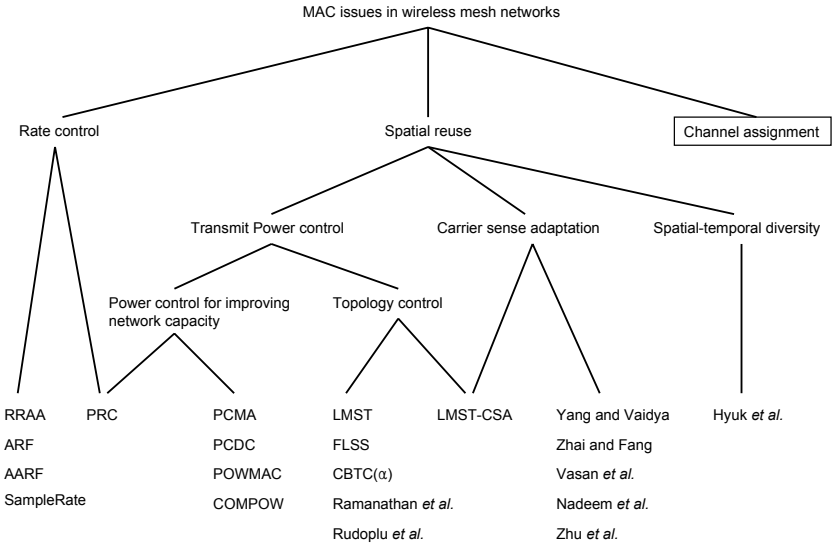



**Fig. 4.2.** A taxonomy of MAC issues in WMNs.

The major function of MAC in WMNs is to arbitrate access to the open and shared medium, with the objective of maximizing network capacity and achieving some level of fairness among users. In particular, there are several PHY/MAC attributes that can be used to improve spatial reuse, mitigate interference and maximize network capacity: (i) the *transmit power* each node uses for communications, (ii) the *carrier sense threshold* each node uses to determine if the shared medium is idle,

(iii) the *channel* on which the node transmits, and (iv) the *time intervals* in which each node gains access to the channel. Note that the carrier sense threshold specifies the received signal strength above which a node determines that the medium is busy and will not attempt for transmission. The first two attributes control the sharing range of the wireless medium in the spatial domain and ultimately the degree of spatial reuse. The third attribute exploits use of non-overlapping channels to mitigate interference. The last attribute leverages temporal/spatial diversity and aims to schedule transmission of packets that may potentially interfere with one another in different time intervals. All these attributes affect the signal-to-interference-plus-noise ratio (SINR) at a receiver. Because the SINR is directly related to the data rate which a transmission can sustain, another PHY/MAC attribute that can be tuned to enhance the overall system performance is the data rate. Fig. 4.2 gives a taxonomy of MAC research issues in WMNs. Note that the issue of channel assignment (which is boxed in Fig. 4.2) is related to the topics to be discussed in Section 4.4, and is further categorized in Fig. 4.5. In what follows, we first outline research issues for each PHY/MAC attribute, and then summarize the state of the art.

*(1) Controlling the Sharing Range of the Wireless Medium and Increasing Spatial Reuse*

One can increase the level of spatial reuse by either reducing the transmit power or increasing the carrier sense threshold (thereby reducing the carrier sense range). The first research issue is how each node determines its transmit power and carrier sense threshold (in a distributed and self-adjusting manner) so that (i) network connectivity is maintained; (ii) the MAC-level interference is mitigated; and (iii) the spatial reuse is utilized (i.e., as many concurrent connections as possible are enabled, subject to maintaining necessary SINR for decoding at certain data rates).

Several interesting related research issues are – What is the relationship between the transmit power and the carrier sense threshold? Will tuning one parameter implies the other? What is the trade-off between (i) increasing the level of spatial reuse by using smaller power or larger carrier sense threshold and (ii) decreasing individual data rates each node can afford (because of the decrease in the SINR as a result of using smaller power/larger carrier sense threshold)? Specifically, when the transmit power decreases, the SINR decreases as a result of the smaller received signal [17, 18]. Similarly, when the carrier sense threshold increases, a node may determine the medium to be idle when some other concurrent transmissions (whose signals received at the node do not exceed $T_{cs}$) are in progress. This leads to the increase in the interference level and the decrease in the SINR. In both cases, the receiver may not be able to correctly decode the signal and the data rate sustained by each transmission may decrease. Finally, what is the minimal information that needs to be exchanged among mesh nodes in order to realize a (sub-)optimal solution (if any)? Answers to these questions are important research issues in order to fully exploit spatial reuse.

*(2) Exploiting Temporal/Spatial Diversity*

Another dimension of improving the network capacity is through *joint temporal and spatial diversity.* Specifically, the overall capacity can be increased by exploiting spatial diversity that exists among a number of multi-hop paths. Packets that are routed along these paths can be scheduled to take place simultaneously if their transmissions do not interfere with each other (significantly). In this manner, even if only single channels are available (e.g., without multi-radios or multi-channels) it is possible that the achievable throughput on a multi-hop wireless path is *only* limited by intra-flow interference.

There are, however, two issues that must be addressed in order to realize spatial diversity. First, the set of paths along which transmissions can take place with the least inter-flow interference must be identified, perhaps with received signal strength measurements. Second, based on the set of non-interfering paths, the order in which packets of different connections are scheduled to be transmitted must be determined, with the objective of mitigating interference.

*(3) Exploiting Availability of Multiple Channels*

Traditional multi-hop wireless networks are mostly comprised of single-radio nodes. Such networks may suffer from capacity degradation due to the half-duplex transmission capability of the wireless medium. A solution is thus to equip nodes with multiple radio interfaces and assigning orthogonal channels to radios. In this manner, nodes can communicate simultaneously with the minimal interference, although they are with the interference range of each other. Even in networks with only single-radio nodes, capacity improvement can be expected by enabling nodes with the interference range of each other to operate on different channels to minimize the amount of interference. Currently the IEEE 802.11b/g and IEEE 802.11a standards provide, respectively, 3 and 12 orthogonal channels which can be used simultaneously within a neighborhood.

A simple design for multi-radio and multi-channel networks would be to equip each node with the same number of radios as the number of orthogonal channels. However, due to both the economical and technical reasons only a limited number of radios may be equipped at each node. The research issue is then how each node determines the channel on which each of its radios will operate, in order to reduce the interference caused by simultaneous transmissions on the same channel. Moreover, channel assignment is usually considered in conjunction with routing (Section 4.4). How to jointly select a route and assign channels to radio interfaces along the route is an important and active research area.

*(4) Exercising Rate Control*

Rate control refers to the process of dynamically adapting the data rate according to the channel status, with the aim of choosing an optimal data rate for the given channel condition. An example is the auto-rate function available in most IEEE 802.11 a/b/g chipsets. There are 4 data rates (1, 2, 5.5, 11 Mb/s) available in 802.11b and 8 data

rates (6, 9, 12, 18, 24, 36, 48, 54 Mb/s) available in 802.11 a/g. Usually the higher the SINR, the higher the data rate. For a given SINR, one may then choose the highest possible data rate (that allows correct decoding) in order to maximize the throughput.

The procedure of rate control consists of two phases: *channel estimation* and *rate selection*. The major research issues to be considered are: (i) which metric should be used to measure the channel quality? and (ii) which design rule associated with the metric should be used to select a new data rate?

### 4.2.1 Transmit Power Control

#### Graph-Model-Based Topology Control

The issue of transmit power control has been extensively studied in the context of topology maintenance by graph-theoretic approaches [19]- [27], where the major objective is to reduce power consumption, mitigate MAC interference, while preserving network connectivity. Since the energy required for transmission increases with the distance (at least in the order of two), it makes sense from the perspective of energy saving to replace one long link with several short links. Furthermore, reducing the transmit power also mitigates MAC interference, which in turn improves the network capacity (as a result of less MAC-level collisions and retransmissions). However, the transmit power cannot be reduced to the extent that network connectivity is not preserved.

A common notion of neighbors adopted in these power control algorithms is that two nodes are considered neighbors and a wireless link exists between them in the corresponding communication graph, if their distance is within the transmission range (as determined by the transmit power, the path loss model, and the receiver sensitivity). Algorithms that adopt this notion are collectively called *graph-model-based topology control*. Under this notion, topology control aims to keep the node degree in the communication graph low, subject to the network connectivity requirement. This is based on the common assertion that a low node degree usually implies low interference.

Rodoplu *et al.* [26] introduced the notion of *relay region* and *enclosure* for the purpose of power control. For any node $i$ that intends to transmit to node $j$, node $j$ is said to lie in the relay region of a third node $r$, if node $i$ will consume less power when it chooses to relay through node $r$ instead of transmitting directly to node $j$. The enclosure of node $i$ is then defined as the union of the complement of relay regions of all the nodes that node $i$ can reach by using its maximal transmission power. It is shown that the network is strongly connected if every node maintains links with the nodes in its enclosure and the resulting topology is a minimum power topology. A two-phase distributed protocol was then devised to find the minimum power topology for a static network. In the first phase, each node $i$ executes local search to find the enclosure graph. This is done by examining neighbor nodes which a node can reach by using its maximal power and keeping only those that do not lie in the relay regions of previously found nodes. In the second phase, each node runs the distributed Bellman-Ford shortest path algorithm upon the enclosure graph, using the

power consumption as the link cost. When a node completes the second phase, it can either start data transmission or enter the sleep mode to conserve power. To deal with limited mobility, each node periodically executes the distributed protocol to find the enclosure graph. This algorithm assumes that there is only one data sink (destination) in the network, which may not hold in practice. Also, an explicit propagation channel model is needed to compute the relay region.

Ramanathan *et al.* [25] presented two centralized algorithms, i.e.,CONNECT and BICONN-AUGMENT, to minimize the maximum power used per node while maintaining the (bi)connectivity of the network. CONNECT is a simple greedy algorithm that iteratively merges different components until only one remains. Augmenting a connected network to a bi-connected network is done by BICONN-AUGMENT, which uses the same idea as in CONNECT to iteratively build the bi-connected network. In addition, a post-processing phase can be applied to ensure per-node minimality by deleting redundant connections. Two distributed heuristics, LINT and LILT, are introduced for mobile networks. In LINT, each node is configured with three parameters - the desired node degree $d_d$, a high threshold $d_h$ on the node degree, and a low threshold $d_l$. Every node will periodically check the number of active neighbors and change its power level accordingly, so that the node degree is kept within the thresholds. LILT further improves LINT by overriding the high threshold when the topology change indicated by the routing update results in undesirable connectivity. Both CONNECT and BICONN-AUGMENT are centralized algorithms that require global information, thus cannot be directly deployed in the case of mobility. On the other hand, the proposed heuristics LINT and LILT cannot guarantee that network connectivity is preserved.

CBTC($\alpha$) [19] is a two-phase algorithm in which each node finds the minimum power $p$ such that transmitting with $p$ ensures that it can reach some node in every cone of degree $\alpha$. The algorithm was analytically shown to preserve network connectivity if $\alpha < 5/6$. It also ensured that every link between nodes is bidirectional. Several optimizations to the basic algorithm are also discussed, which include: (i) a shrink-back operation can be applied to allow a boundary node to broadcast with less power, if doing so does not reduce the cone coverage; (ii) if $\alpha < 2/3$, asymmetric edges can be removed while maintaining network connectivity; and (iii) if there exists an edge from $u$ to $v_1$ and from $u$ to $v_2$, respectively, the longer edge can be removed while preserving connectivity, as long as $d(v_1, v_2) < \max(d(u, v_1), d(u, v_2))$. An event-driven strategy was proposed to reconfigure the network topology in the case of mobility. Each node is notified when any neighbor leaves/joins the neighborhood and/or the angle changes. The mechanism used to realize this requires state to be kept at, and message exchanges among neighboring nodes. The node then determines whether it needs to rerun the topology control algorithm.

Li and Hou [24] proposed a topology control algorithm, called *Local Minimum Spanning Tree (LMST)*, for multi-hop wireless networks with limited mobility. The topology is induced by having each node build its local MST independently (with the use of information locally collected) and only keep one-hop on-tree nodes as neighbors. Specifically, LMST is composed of three phases: *information collection*, *topol-*

*ogy construction*, and *determination of transmit power*, and an optional optimization phase: construction of topology with only bidirectional edges. In the information exchange phase, the information needed by each node for topology construction is obtained by having each node broadcast periodically a Hello message using its maximal transmit power. A Hello message should at least include the node id and the position of the node. In the topology construction phase, each node independently applies Prim's algorithm [28] to obtain its local minimum spanning tree. Then, by measuring the received signal strength of a Hello message, each node determines the specific power level necessary to reach each of its neighbors in the phase of determining transmit power.

LMST can further optimize the topology by replacing all the uni-directional links with bi-directional ones. To this end, every node may probe each of its neighbors to find out whether or not the corresponding edge is uni-directional, and in the case of a uni-directional edge, either deletes the edge or notifies its neighbor to add the reverse edge. The capability of forming a topology that consists of only bi-directional links is important for link level acknowledgments, and critical for packet transmissions and retransmissions over the unreliable wireless medium. It has been proved that LMST possesses several desirable properties: (i) the topology constructed under LMST preserves network connectivity; and (ii) the degree of any node in the resulting topology is bounded by 6. LMST has been also extended to include heterogeneous networks [21, 22], where the maximum transmit power can be different for each node, and to maintain $k$-connectivity ($k \geq 2$) [20, 23].

In spite of all the efforts in deriving all the graph theoretically grounded results, the underlying assertion that a low node degree usually implies low interference does not actually hold under the physical *Signal-to-Interference-Noise-Ratio* (SINR) model. As discussed in [29–31], this is because under the physical model, whether the interference — the sum of all the signals of concurrent, competing transmissions received at the receiver — affects the transmission activity of interest depends on the SINR at the receiver, which in turn depends on the transmit power of all the transmitters and their relative positions to the receiver of interest. The node degree under the graph model, however, does not adequately capture interference. In particular, a transmission of interest may fail because another concurrent transmission causes the SINR at the receiver to fall below the minimal SINR required for the receiver to decode the symbols correctly. This could occur even if the competing transmitter is outside the transmission range of the receiver. There are two undesirable consequences as a result of the inadequacy of graph-model-based topology control under the physical model. First, because the node degree does not capture interference adequately, the interference in the resulting topology may be high, rendering low network capacity. Second, a wireless link that exists in the communication graph may not in practice exist under the physical model, because of high interference (and consequently low SINR). As a result, the network connectivity may not even be sustained.

**Power Control for Improving Network Capacity**

Use of transmit power control for maximizing network capacity has been considered in [32]-ch04-Muqattash:04. Monks *et al.* [33] proposed the *Power Controlled Multiple Access (PCMA)* algorithm, in which the receiver advertises its interference margin that it can tolerate on an out-of-band channel and the transmitter selects its power in order not to disrupt any ongoing transmissions. Similar to IEEE 802.11 RTS/CTS handshake, PCMA uses RPTS/APTS handshake to decide the minimal transmission power for successful frame reception. PCMA further introduces an additional channel, i.e., the busy tone channel in order to implement the noise tolerance advertisement. Any transmitter must sense the busy tone to decide its transmit power for a minimum time period. As compared to the IEEE 802.11 protocol, PCMA can improve the throughput by more than a factor of two in high-density networks.

Muqattash and Krunz proposed a similar power control protocol, called *Power Controlled Dual Channel (PCDC)* [34]. The PCDC protocol constructs the network topology by overhearing RTS/CTS packets, and the computed interference margin is announced on an out-of-band channel. The basic idea of PCDC is to employ a distributed algorithm for computing a minimal connectivity set (i.e., a minimum set of nodes that guarantees connectivity of the node to the network) in order to find the lowest possible power level while preserving the network connectivity and proper MAC functions. As compared to the IEEE 802.11 standard, PCDC can achieve improvements of up to 240% in channel utilization and over 60% in end-to-end throughput, and a reduction of more than 50% in energy consumption. However, it should be noted that the adaptive computing process for the connectivity set may require extensive computing overhead at each node. Muqattash and Krunz also proposed a single channel protocol called *POWMAC* [35] for exchanging the interference margin information.

Narayanaswamy *et al.* [36] developed a power control protocol, called *COMPOW*. The authors argued that if each node uses the smallest common power required to maintain network connectivity, the traffic carrying capacity of the entire network is maximized, the battery life is extended, and the contention at the MAC layer is reduced. In COMPOW each node runs several routing daemons in parallel, one for each power level. Each routing daemon maintains its own routing table by exchanging control messages at the specified power level. By comparing the entries in different routing tables, each node can determine the smallest common power that ensures the maximal number of nodes are connected. Specifically, let $N(P_i)$ denote the number of entries in the routing table corresponding to the power level $P_i$. Then the adequate power level for data packets is simply set to the smallest power level $P_i$ for which $N(P_i) = N(P_{\max})$. The major drawback of COMPOW is its significant message overhead, since each node runs multiple daemons, each of which has to exchange link state information with the counterparts at other nodes. COMPOW also tends to use higher power in the case of unevenly distributed nodes. Finally, since the common power is collaboratively determined by the all nodes inside the network, global reconfiguration is required in the case of node joining/leaving.

### 4.2.2 Adaptation of Carrier Sense Threshold

Recently a number of studies have focused on exploiting IEEE 802.11 physical carrier sense to increase the level of spatial reuse [37]- [40]. By physical carrier sense, it means that before attempting for transmission, a node senses the medium and defers its transmission if the channel is sensed busy, i.e., the strength of the received signal exceeds a certain threshold $CS_{th}$. Carrier sense reduces the likelihood of collision by preventing nodes in the vicinity of each other from transmitting simultaneously, while allowing nodes that are separated by a safe margin (termed as the carrier sense range) to engage in concurrent transmissions.

Given a predetermined transmission rate, Zhu *et al.* [39] derived a simple condition for the carrier sense threshold, in order to cover the entire interference range for several regular topologies. Zhu *et al.* also proposed in [40] a dynamic algorithm for adjusting the carrier sense threshold to exploit spatial reuse. The algorithm calculates, based on the estimate of the current local interference condition, a near-optimal value for the carrier sense threshold. In this manner, the SINR at each node can be kept above the desired threshold by local measurement and information exchange. However, the proposed feedback algorithm is essentially heuristic based. Thus, the challenge remains on how to design a theoretically grounded, self-adapting algorithm for tuning the carrier sense threshold, with the aim of improving the network capacity.

Vasan *et al.* [38] proposed an algorithm, entitled *echos*, for on-line tuning of the carrier sense threshold in order to allow more flows to co-exist in IEEE 802.11-based hotspot wireless networks. Nadeem *et al.* [37] proposed a location-enhanced DCF algorithm that exploits location information to exploit spatial reuse for given transmission rates.

Yang and Vaidya [41] are perhaps the first to address the impact of physical carrier sense on Shannon capacity of single-rate, multi-hop wireless hoc networks, while *taking into account the MAC layer overhead*. Under the assumption of a dense network, they derived an analytical model that characterizes the relationship between the Shannon capacity and the carrier sense range. Note that they only considered first-tier interference in the calculation of SINR. Based on the derived model, they made the following key observations: (i) the MAC overhead has a fundamental impact on the selection of the optimal carrier sense threshold. By selecting a larger value of the carrier sense threshold, both the bandwidth-independent MAC overhead and the bandwidth-dependent MAC overhead can be reduced, which in turn, improves the utilization of each individual wireless link; and (ii) the optimal value of the carrier sense threshold depends on the level of channel contention, packet size, and other factors affecting the bandwidth-dependent and bandwidth-independent overheads. With the use of an inappropriate carrier sense threshold, the aggregate network throughput may severely degrade.

Zeng and Hou [42] analyzed IEEE 802.11 DCF in single-rate, multi-hop wireless networks with consideration of the effects of physical carrier sense, SINR, and collision caused by accumulative interference. Specifically, they substantially extended Calì's analytic model [43] and rigorously modeled, with these effects considered,

channel activities governed by IEEE 802.11 DCF in multi-hop wireless networks. They showed that as in WLANs, the choice of the contention window size can greatly impact the system throughput in multi-hop wireless networks. However, the optimal value of the contention window size is much smaller. This is because in multi-hop wireless networks, (i) physical carrier sense has already, to some extent, restricted nodes from accessing the medium, and (ii) a node may be silenced not only by transmissions in its vicinity, but also by accumulative interference that exceeds the carrier sense threshold. While a larger attempt probability increases the collision probability, it also helps to reduce the idle periods before successful transmissions, and mitigate the above effects. Moreover, given the minimal SINR threshold, the optimal carrier sense range is smaller than the conventional value used (provided that the contention window size is tuned accordingly). This suggests that, as long as the contention window size is appropriately controlled, the systems throughput can be further improved by allowing more concurrent transmissions and increasing spatial reuse.

Zhai and Fang [44] investigated the impact of physical carrier sense in multi-rate, multi-hop wireless networks where nodes have different levels of transmit power. They also considered the impacts of SINR, node topology, hidden/exposed nodes, and bidirectional handshakes to determine the optimal carrier sense range for maximizing the throughput. Through analysis and simulation, they made the following observation: (i) the optimal carrier sense threshold for one-hop flows does not seem to work well for multi-hop flows. This implies that characteristics unique to multi-hop flows should be carefully considered to find the optimal carrier sense threshold; this observation is consistent with that in [42]; (ii) the optimal carrier sense threshold derived for different data rates is similar to each other. This suggests that a single value of the carrier sense threshold can be used for different data rates; and (iii) without use of an adequate carrier sense threshold, higher data rate does not necessarily give higher throughput.

### 4.2.3 Joint Control of Transmit Power and Carrier Sense Threshold

Fuemmeler *et al.* [45] studied the relation between the transmit power and the carrier sense threshold in determining the network capacity. They concluded that transmitters should keep the product of their transmit power and carrier sense threshold fixed at a constant, i.e., the lower the transmit power, the higher the carrier sense threshold (and hence the smaller the carrier sense range), and vice versa. A combination of lower transmit power and higher carrier sense leads to a large number of concurrent transmissions, with each transmission sustaining a small data rate. On the other hand, a combination of higher transmit power and lower carrier sense threshold leads to a small number of concurrent transmissions, with each transmission sustaining a large data rate. Although the analysis gives a general trend, it does not give guidelines on how to select the two parameters to *maximize* the network capacity.

Kim *et al.* [46] studied the relationship between physical carrier sense and Shannon capacity, and showed that (i) in the case that the achievable channel rate follows the Shannon capacity, spatial reuse depends only on the ratio of the transmit power to the carrier sense threshold; and (ii) in the case that only a set of discrete data rates

are available, tuning the transmit power offers several advantages that tuning the carrier sense threshold cannot, provided that there is a sufficient number of power levels available. Point (i) implies that, to improve (or in the best case optimize) network capacity, one can tune one parameter, while fixing the other at an appropriate value.

Yang *et al.* [47] extended both Bianchi's model [48] and Kumar's model [49], and characterized the channel activities governed by IEEE 802.11 DCF in single-rate, multi-hop wireless networks from the perspective of an individual sender. In particular, they incorporated the effect of PHY/MAC attributes, such as transmit power and physical carrier sense, that need not be considered in WLANs but become extraordinarily important in multi-hop wireless networks, and derive the throughput attained by each sender. With the use of the analytical model derived, they then investigated the impact of transmit power and carrier sense threshold on network capacity, and identified a simple operating condition under which the network may attain throughput that is close to its optimal value. Specifically, they found that high system throughput can be achieved when the area within the carrier sense range silenced by a sender $s$ is reduced as much as possible *under the premise that it still covers the interference area of its intended receiver* $r$. This increases spatial reuse while not deteriorating collisions due to the hidden node problem.

Based on the insight shed from the above analytical model, Yang *et al.* [47] proposed a distributed and localized algorithm, called *Local Minimum Spanning Tree with Carrier Sense Adjustment (LMST-CSA)* that determines both the transmit power and the carrier sense threshold of a node. In LMST-CSA, each node determines its transmit power based on LMST [24], and then controls its carrier sense threshold so that the desirable operating condition is met. Simulation results show that *LMST-CSA* achieves higher throughput as compared to conventional IEEE 802.11 DCF, *LMST with no carrier sense adjustment*, and *LMST with static carrier sense adjustment*.

### 4.2.4 Exploitation of Spatial-Temporal Diversity

The problem of mitigating interference and improving network capacity was also considered from the angle of spatial-temporal diversity in [32]. Lim *et al.* focused on transporting downstream traffic at gateway nodes with Internet access and proposed to construct, based on the received signal strengths (RSS) measurements, a virtual coordinate system. This is in contrast to most existing work which relies on geographic locations of wireless mesh nodes. The reason for using RSS measurements, rather than geographic distances, among neighbors as the references is because RSS measurements are more "representative" in determining the level of interferences between nodes. Specifically, the RSS measurements between a node $\mathbf{n}$ and its neighbors are represented by the $p \times p$ square matrix $\mathbf{S}$, the columns of which can be considered as the coordinates of the corresponding nodes in a $p$-dimension space. Note that the $i$th column vector of $\mathbf{S}$ is the RSSs measured by the $i$th node from all the nodes. As these coordinates are correlated with each other, it is difficult to identify components that play an important role in determining the interferences. Hence Lim *et al.* constructed an orthogonal virtual coordinate system with a smaller dimensionality by using singular value decomposition, and used the "virtual distance" between mesh

nodes to infer the level of interferences between them. With the use of the coordinate system, they were able to determine the sets of paths along which transmissions can take place with the least inter-flow interference.

Based on the sets of of non-interfering paths, a gateway node then determines the order in which a gateway node schedules frames of different connections to be transmitted. To allow a gateway node to send frames consecutively in an non-interruptible manner, we leverage the *transmission opportunity (TXOP)* option in the IEEE 802.11e specification [50]. That is, a gateway node that succeeds in grasping the medium is granted the right to use the medium for a period of time specified by TXOP. The gateway uses a TXOP to transmit multiple frames, with SIFS (instead of DIFS) as the inter-frame space *between* the sequence of DATA-ACK exchanges. If the DATA-ACK exchange has been completed, and there is still time remaining in the TXOP, the node may transmit another frame (after an idle time of SIFS), provided that the frame to be transmitted and its necessary acknowledgment can fit into the time remaining in the TXOP. The experimental results showed that the downstream throughput of a gateway node in a wireless mesh network can be improved by 10 - 35% under various network topologies and traffic distributions. Also, the proposed approach requires only minimal code change in the gateway nodes and does not require any extra hardware.

### 4.2.5 Exploitation of Channel Diversity Through Channel Assignment

*Multi-channel MAC (MMAC)*

The MMAC protocol [51] is motivated by the fact that most of the existing MAC protocols are designed for single-channel operations, although the IEEE 802.11 standard supports the use of multiple channels. Under MMAC, each node is equipped with only one transceiver, but can switch channels dynamically with the objective of mitigating interference and improving network capacity. To support dynamic negotiation of channels, the time is divided into fixed-time intervals using beacons, and a small window called the *ATIM window* at the beginning of each interval is used to negotiate channels for transmitting packets. In an ATIM window, all nodes listen to a pre-defined, default channel on which beacons and ATIM packets are transmitted. One important information in an ATIM packet is the *preferable channel list (PCL)* that indicates which channel is preferred for the node. PCL is maintained at both the source and the destination.

When a node receives an ATIM packet, it selects a channel and sends to the sender an ATIM-ACK packet that includes the selected channel. The channel to be selected is determined from (i) the information included in the PCL sent by the sender and (ii) the PCL locally kept. Also, the number of source-destination pairs that have selected a channel is counted by overhearing ATIM-ACK and ATIM-RES packets. The selection procedure that a node uses then attempts to balance the channel load as much as possible so that the bandwidth waste caused by contention and backoff is reduced. As the simulation study indicated, MMAC improves network throughput significantly, especially when the network is highly congested. This is, in part, due

to the fact that MMAC successfully exploits multiple channels to achieve higher throughput than IEEE 802.11 DCF.

*Asynchronous Multichannel Coordination Protocol (AMCP)*

AMCP is a distributed medium access protocol that utilizes multiple channels to address *starvation* in a multi-hop wireless network [52]. It is argued that a single-channel CSMA system may suffer from starvation when CSMA based access is used in a multi-hop environment. If the senders of two contending flows are not within the carrier sensing range of each other and have an asymmetric view of the channel state, then one transmitter may achieve significantly higher throughput than the other. This is because this transmitter does not experience collision, but the other suffers from RTS failures and exponential back-off. The starvation problem that thus arises is called *Information Asymmetry (IA)* problem. The other source of starvation is caused by the so-called *Flow-in-the-Middle (FIM)* problem. This problem occurs when the transmitter of a flow has neighboring transmitters which are not within the carrier sense range of each other. In this case, the middle flow can barely have any transmission opportunity since its transmission activities is deferred by neighboring nodes. Both IA and FIM problems are caused by the asymmetry of multi-hop topology and the use of carrier sense.

To cope with starvation, one simple approach is to keep separate channels for control and data transmission. This alleviates starvation since contentions only occur on the control channel for transmission of control packets whose length is comparable to the back-off period. Following this approach, AMCP designates a control channel for nodes to contend and reserve data channels by exchanging RTS/CTS packets according to 802.11 DCF. Once a control packet is exchanged successfully, both the sender and the receiver switch to the reserved data channel, and transmit a data packet. After a data packet is successfully transmitted on the reserved channel, the sender and receiver return to the control channel and set all channels as unavailable for a pre-determined time interval except the one just used. They may contend for the reserved data channel immediately or contend for other data channels after the specified time interval elapses. The simulation study showed that AMCP not only utilizes multiple channels to achieve a significant aggregate throughput gain (as compared to single-channel systems), but also adequately addresses the starvation problem.

*MAXchop*

Mishra *et al.* [53] addressed the fairness issue in IEEE 802.11 hotspot networks from the perspective of channel assignment. In an uncoordinated environment of hotspot access points, proper channel assignment is critical. The APs that implement channel assignment algorithms have to ensure that the total wireless bandwidth is divided fairly among interfering hotspot APs. No hotspot should have a higher priority on the total bandwidth over others, irrespective of the number of clients. Providing proportional fairness in this environment will require additional coordination between APs/clients across different management domains.

The MAXchop algorithm was proposed to utilize channel hopping to improve fairness of any existing channel assignment. In MAXchop, in each slot the APs utilize a specific channel assignment that may have been computed using existing distributed algorithms. In different slots, the APs utilize different channel assignments. The channel assignment used in a slot is different from that in the previous slot, but is yet locally the best. MAXchop enables the APs to utilize all the channel assignments so as to uniformly divide the available bandwidth among the APs. Consequently, this approach ensures that the long term throughput that each AP attains is averaged over multiple different channel assignments. Experiments results showed that with partially overlapped channels as well as with non-overlapped channels, MAXchop improves both fairness and throughput.

*Component Level Channel Assignment (CLCA)*

Practical considerations such as the switching delay and the synchronization and scheduling overheads greatly impact the performance of channel assignment. Vedantham *et al.* [54] utilized the concept of connected component as granularity of assignment to mitigate such overheads. This is in contrast to existing channel assignment algorithms which assign channels to packets, links, or flows. A connected component in a flow graph is the largest subgraph, such that there exists a path between any node to all the other nodes in the subgraph. Besides its simplicity, this algorithm has the following advantages: (i) there is no need to change the off-the-shelf radio hardware or MAC algorithms, and (ii) there is no synchronization requirement, channel scheduling overheads, or switching between channels to serve data flows.

Conceptually, this algorithm involves assigning a single channel to all the nodes which are included in a component (formed by nodes which make mutually intersecting flows). All inks in a connected component induced by the underlying flow graph operate in a single channel. However, different connected components can potentially operate on different channels. The algorithm has two phases: *path selection* and *channel assignment*. In the path selection phase, paths that minimize the number of intersections in the network are selected and, components are built up with the selected paths. Once the component set has been determined, channels are assigned to the components obtained in the first phase. In the channel assignment phase, channels are so assigned that the contention between different components in the underlying flow graph is minimized.

*Slotted Seeded Channel Hopping (SSCH)*

Bahl *et al.* [55] proposed the *Slotted Seed Channel Hopping (SSCH)* protocol in which nodes with a single interface are allowed to switch across channels in such a way that nodes desiring to communicate overlap, while disjoint communications mostly do not overlap (and hence do not interfere) with each other. In SSCH, the time allocated to a single channel is defined as a slot, which is 10 ms in the implementation and corresponds to 35 packet transmission times at 54 Mbps. In SSCH, each device picks multiple (e.g., 4) sequences, each of which is uniquely determined by the seed of a pseudo-random generator, and follows them in a time-multiplexed

manner. When device $A$ would like to talk to device $B$, it waits until it is on the same channel as $B$. If device $A$ would like to talk to device $B$ frequently, it adopts one or more of device $B$'s sequences, thereby increasing the time they are on the same channel.

For the channel hopping mechanism to work, the sender learns the current sequences the receiver uses, via a seed broadcast mechanism. Every node broadcasts its channel schedule in each slot so that nodes can know each other's channel hopping schedule. This is termed as *optimistic synchronization*. Schedules are updated in two ways: each node will loosely synchronize the slot's start and finish time with other nodes, or it will overlap another node's schedule if it is going to send packets to this node. Also, the node will delay channel switching when it is communicating with another node until it finishes. Another strategy called *partial synchronization* is used for assigning channels, changing schedule and preventing channel congestion from taking place. Simulation results showed that both in the single-hop and multi-hop cases, SSCH performs significantly better than IEEE 802.11a achieving significant capacity improvement.

### 4.2.6 Rate Control

Rate control algorithms have been studied quite extensively in [56]- [65], and some of them have also been implemented in real products [56, 60]. As mentioned above, rate control aims to adjust the channel data rate with respect to the time-varying channel status. In principle, rate control consists of two phases: *channel estimation* and *rate selection*. The most commonly used metrics for estimating the quality of a channel include probe packets [56, 60, 62], consecutive successes/losses [57, 60, 62], and SINR [57, 59, 65]. The commonly used design rules for selecting a new data rate are increasing/decreasing the data rate on consecutive transmission successes (packet losses) and exploiting probe packets to assess new rates.

Wong *el al.* [66] conducted a study on challenges of rate control and explored a new design space. They evaluated several critical design guidelines that have been followed by most of existing algorithms: (i) decreasing the data rate on severe packet loss, (ii) using probe packets to assess the new rate, (iii) using consecutive success/failure as the index to increase/decrease the data rate, (iv) using PHY metrics such as SINR to infer the new data rate. Their experiments surprisingly showed that the above guidelines can be quite misleading, and may result in severe throughput degradation of up to 70%. They then proposed a *Robust Rate Adaptation Algorithm (RRAA)* based on the following ideas. First, they used the short-term loss ratio to opportunistically guide the rate selection. Second, they leveraged the per-frame RTS option, and used an adaptive RTS filter to prevent collision losses. They showed that the throughput of RRAA can be improved up to 143% in realistic field trials, as compared to the well known algorithms such as ARF, AARF, and SampleRate.

Kim *et al.* [46] proposed a joint power and rate control algorithm from the perspective of maximizing the level of spatial reuse. Following their observation that spatial reuse depends only on the ratio of the transmit power to the carrier sense threshold (Section 4.2.3), they proposed to tune the transmit power, while keeping

the carrier sense threshold fixed at an appropriate value. Then, they devised a localized power and rate control algorithm, called *Power and Rate Control (PRC)*, which enables each transmitter to adaptively perceive and determine its transmit power and data rate. The transmit power is so determined that the transmitter can sustain the highest possible data rate, while keeping the adverse interference effect on the other neighboring concurrent transmissions minimal. Simulation studies showed that, as compared to existing tuning algorithms for the carrier sense threshold, PRC improves the network capacity for up to 22%.

## 4.3 Example Device Driver Support for Cross Layer Design and Optimization

As discussed in Section 4.1, the traditional notion of a link is no longer well-defined in wireless environments, because characteristics of wireless links are now determined by several PHY/MAC control knobs, as well as inter-flow interference, multipath fading, temperature and humidity variations, and/or the presence of obstacles in the communication path. This implies that in order to optimize the network performance, PHY/MAC attributes should be exported to higher layer protocols in order to enable cross layer design and optimization.

In addition to the above technical concerns, the lack of an open, modular programming environment also imposes a hindrance to the wide deployment of cross-layer design/optimization algorithms in WMNs. Although many of the previous efforts have made their source code available [3, 4, 6, 8, 9], the software (such as customized device drivers, address resolution modules, routing daemons, and name servers) is often implemented in an ad-hoc manner, lacks in structural modularity, and does not come with well-defined APIs for experimentation and performance tuning. This presents a major hurdle for networking researchers to neatly incorporate their research results in the most performance-efficient manner, and empirically assess the algorithm/protocol performance.

As an example programming environment, we introduce in this section the *Transparent Device Driver layer (TDD)* proposed by Kung *et al.* in [67] and situated above the IEEE 802.11 device firmware. As part of the CUWiN software, TDD leverages the Atheros chipset, and the open-source Madwifi driver [68] in Linux and similar device drivers in NetBSD.[1] Although commodity 802.11 interfaces typically partition the MAC functionalities between hardware/firmware on the card and the software driver running in the kernel, the Atheros chipset does not require the loading of firmware. The chipset instead relies on a *Hardware Access Layer (HAL)* module provided in the binary form only. The HAL module operates between the hardware and the device driver to manage many of the chip-specific operations and to enforce required FCC regulations. It is similar to firmware, in that it prevents users from setting invalid operating parameters, but implements fewer 802.11 functionalities than

[1] Note that the Madwifi driver for Linux was originally derived from NetBSD.

other firmware. More desirably, it provides an interface for changing various device parameters, including the minimum and maximum contention windows.

### 4.3.1 Architecture and Major Components

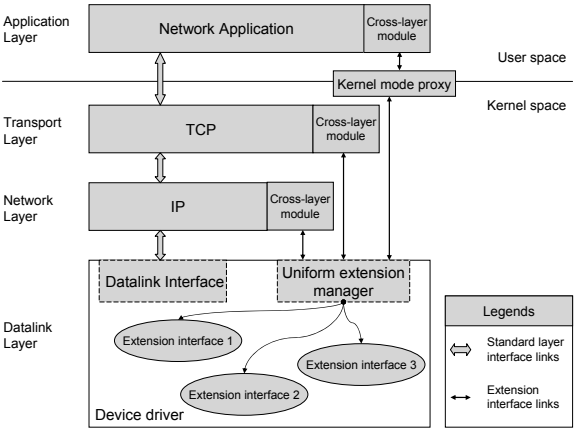



**Fig. 4.3.** The architecture of the uniform extension framework.

Fig. 4.3 shows the architecture of the transparent device driver (TDD). Different from the traditional layered approach, an extension-enabled device driver exports PHY/MAC parameters and events to higher-layer protocol modules. There are three major components in the TDD:

Extension-enabled device driver: The device driver has been extended to export a set of PHY/MAC attributes and events in the form of *extension specification*. The specification serves as a service agreement between the device driver and a higher-layer protocol module that uses it. To implements an extension, a device driver implements the get/set handlers of the PHY/MAC parameters. It also define events, provide the event information to the uniform extension manager, and notify the manager upon occurrence of events.

Cross-layer control module: A cross-layer control module implements a cross-layer design/optimization algorithm. As a client to the uniform extension manager, it registers itself with the uniform extension manager in order to use the facilities provided by the extension-enabled device driver. Through a generic interface, a control module can read and write PHY/MAC parameters exported by the driver. Also, it can subscribe to events of interest defined in an extension specification and provide the corresponding callback functions.

Uniform extension manager: The uniform extension manager is the major component of the TDD. We will elaborate on its internals in Section 4.3.2. Conceptually, it is responsible for (i) loading and unloading extensions, (ii) providing an API for cross-layer control modules to register events of interest and callback functions; (iii) allowing control modules to set/get PHY/MAC parameters via handlers registered by extensions; (iv) maintaining event definition and subscription; and (v) dispatching events to subscribing control modules.

Kernel mode proxy: For user-space programs to gain access to the TDD in the kernel, we introduce a *kernel mode proxy* that serves as a "bridge" between the two entities. Each uniform extension function exported is assigned an unique system call number. The kernel mode proxy is responsible for translating a TDD-related system call and invoking the corresponding uniform extension function, and (ii) delivering events to the handler in the user space.

### 4.3.2 Internals of Uniform Extension Manager

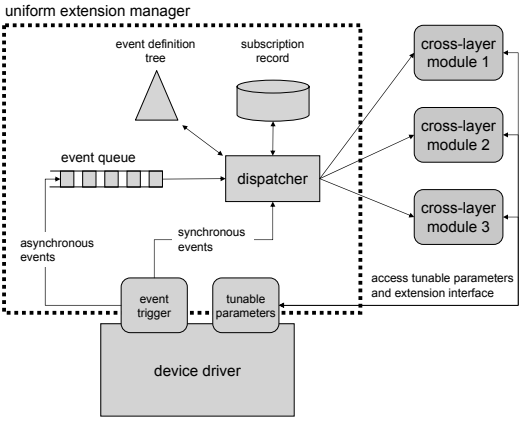



**Fig. 4.4.** Uniform extension manager and event delivery.

Fig. 4.4 shows the internals of the uniform extension manager and the data path in the event delivery mechanism. Table 4.3.2 lists the APIs exported by the uniform extension manager. The uniform extension manager maintains (i) the definition record of all the supported events in an event definition tree; and (ii) the list of subscribers of each event. A cross-layer control module (un-)subscribes to an event with a callback function by calling `AddEventHandler()` (`RemoveEventHandler()`). A device driver generates and delivers an event to the uniform extension manager (and subsequently cross-layer control modules that are interested in the event) by calling `TriggerEvent()`.

**Table 4.1.** The APIs defined in the uniform extension manager.

| Category | Function name | Function description |
|---|---|---|
| Extension management | RegisterExtension()<br>UnregisterExtInterface() | Register/unregister an extension interface module. |
| | FindExtension() | Query whether an extension interface identified by a unique name or id exists. Return a handle to the interface if it exists. |
| Register parameter set/get handlers | RegisterSetHandler()<br>UnregisterSetHandler() | Register or unregister a set handler to the uniform extension manager. |
| | RegisterGetHandler()<br>UnregisterGetHandler() | Register or unregister a get handler to the uniform extension manager. |
| Access to extension parameters | GetExtParam() | Get the value of an extension parameter by invoking the registered get handler. |
| | SetExtParam() | Set the value of an extension parameter by invoking the registered set handler. |
| Event Subscription and delivery | TriggerEvent() | Generate an event and deliver it to the subscribers. |
| | AddEventHandler()<br>RemoveEventHandler() | Subscribe to an event with a callback handler function. |

Depending on the type of events, there are two possible paths for delivering an event to the manager. A *synchronous event* is an event for which the device driver requires feedback from its subscribers. When a synchronous event is triggered, it is delivered by the dispatcher immediately and the device driver that triggers the event waits until all the subscriber handlers are finished. An example of a synchronous event is a *transmit query*, in which prior to the transmission of a frame, the device driver may query the cross-layer control modules for recommendations on the transmit power, the channel on which the frame will be transmitted, or the data rate at which the frame will be transmitted. This facilitates realization of, for example, per-packet power control. Synchronous events make it possible for cross-layer control modules to make decisions upon occurrence of certain events. An *asynchronous event*, on the other hand, is a *notification* message sent by the device driver to the subscriber(s) of that event. Upon reception of an asynchronous event, the event trigger inserts the event into the event queue and wakes up the dispatcher. The dispatcher then delivers the event to the corresponding callback functions.

One point worthy of mentioning is how `TriggerEvent()` is implemented for asynchronous events. As many of the events are triggered by interrupts, `TriggerEvent()` is likely to be invoked by an interrupt handler. However, it is not safe to deliver events inside the context of interrupt handlers, since if for any reason the operation is delayed and the interrupt handler cannot finish, it may interfere the normal system operation. Therefore, we split the task into event creation and event delivery. The `TriggerEvent()` function only creates and puts the event into event queue. A separate kernel thread is created for the dispatcher. The dispatcher thread constantly monitors the event queue and is awakened only when there is a new event. In this manner, the overhead incurred in interrupt handlers is greatly reduced.

### 4.3.3 Desirable Features

The TDD has the following salient features:

**Controlled transparency**: The TDD provides a transparent and generic interface for higher-layer protocol modules to access, through well-defined APIs, a rich set of PHY/MAC attributes and functionalities in the device driver. Specifically, the following PHY/MAC attributes are available: (i) the transmit power level, (ii) the carrier sense threshold, (iii) the data rate, (iv) the receive signal strength index (RSSI), and (v) the channel used to transmit a frame/upon which a frame is received, and (v) the time instant at which a frame is scheduled for transmission/receive. (Note that to obtain the received signal strength, the driver has to instrument the HAL to query, upon receipt of a frame, the value of a specific hardware register). Through an event subscription mechanism, higher-layer protocol modules can also receive timely update of channel status, without directly inserting callback functions in various places of the device driver.

**Flexibility**: The design philosophy of the TDD (and at heart the uniform extension manager) is to provide minimum but crucial functionalities that enable implementation of complicated cross-layer design/control algorithms. The event subscription mechanism is simple, elegant, and allows *multiple* higher-layer protocol modules to subscribe, and be alerted of, PHY/MAC events of interest. They can also register with the event subscription mechanism their callback functions, allowing adequate actions to be taken upon event occurrence. Moreover, the TDD allows the time granularity at which PHY/MAC properties are controlled to be on a *per-packet* or *per-connection* basis, or *permanently* (i.e., until the property is reset).

**Easy Integration and Portability**: Existing upper-layer protocol modules (e.g., routing daemons) can be extended to subscribe events of interest (e.g., frame reception status upon frame arrival), and figure in the information in their decision making. Through dynamic module loading and extension registration, an upper-layer protocol can realize cross-layer optimization if an extension has been implemented, and it falls back to the normal operation if the required extension is not supported by the TDD. This ensures portability.

## 4.4 Routing That Leverages PHY/MAC Attributes in WMNs

As discussed in Section 4.3, with the availability of a modular programming environment that exports PHY/MAC attributes and events to higher-layer protocols, numerous cross-layer design and optimization algorithms/protocols can be designed, implemented and experimented. In this section, we use routing as an example to demonstrate how higher-layer protocols can take advantage of PHY/MAC attributes (such as the channel status and the availability of multiple channels) for optimizing the performance.

Routing in ad hoc wireless networks has been an active area of research for many years. Much of the work in the area was motivated by the need to consider

energy constraints imposed by battery-powered nodes and to deal with node mobility. The research focus is thus to provide routes that are *resilient* to topology change in an *energy-efficient* manner. Unlike ad hoc wireless networks, most of the nodes in WMNs are stationary and thus dynamic topology changes are less of a concern. Also, wireless nodes in WMNs are mostly access points and Internet gateways and thus are not subject to energy constraints. As a result, the focus is shifted from maintaining network connectivity in an energy efficient manner to finding high-throughput routes between nodes, so as to provide users with the maximal end-to-end throughput. In particular, because multiple flows initiated by multiple nodes may engage in transmission at the same time, how to locate routes that give the minimal possible interference is a major issue.

The issue of locating interference-free (or interference-mitigated) routes has been addressed in the literature with roughly two complimentary approaches. First, some of the PHY/MAC attributes have been utilized to define better route metrics that yield high-throughput routes. Note that the conventional route metric is the hop count [69]-[71], and has been used in on-demand, ad-hoc routing protocols such as Ad-hoc On-demand Distance Vector (AODV) and Dynamic Source Routing (DSR). Use of this metric renders routes that are composed of long links. Due to the path loss effect over the distance, these long links are lossy and of low throughput [72]. The performance of routing protocols can be improved by better defining route metrics and explicitly taking into account the quality of wireless links. Second, each wireless node is usually equipped with one or more radios that can be switched among multiple non-overlapping channels. Use of multi-radios and multi-channels has thus been explored to construct interference-free/mitigated routes on which different channels are associated with different radios in order to eliminate intra- and inter-flow interference. The latter approach has been referred to as joint routing and channel assignment. In what follows, we first discuss several route metrics which have been proposed in the literature, and then summarize the various routing protocols with the taxonomy given in Fig. 4.5 as the roadmap.

### 4.4.1 Routing Metrics

*Expected Transmission Count (ETX)*

This metric calculates the expected number of transmissions (including retransmissions) needed to send a frame over a link, by measuring the forward and reverse delivery ratios between a pair of neighboring nodes [72]. To measure the delivery ratios, each node periodically broadcasts a dedicated *link probe* packet of a fixed size. The probe packet contains the number of probes received from each neighboring node during the last period. Based on these probes, a node can calculate the delivery ratio of probes on the link to and from each of its neighbors. The expected number of transmissions is then calculated as

$$ETX = \frac{1}{d_f \times d_r}$$

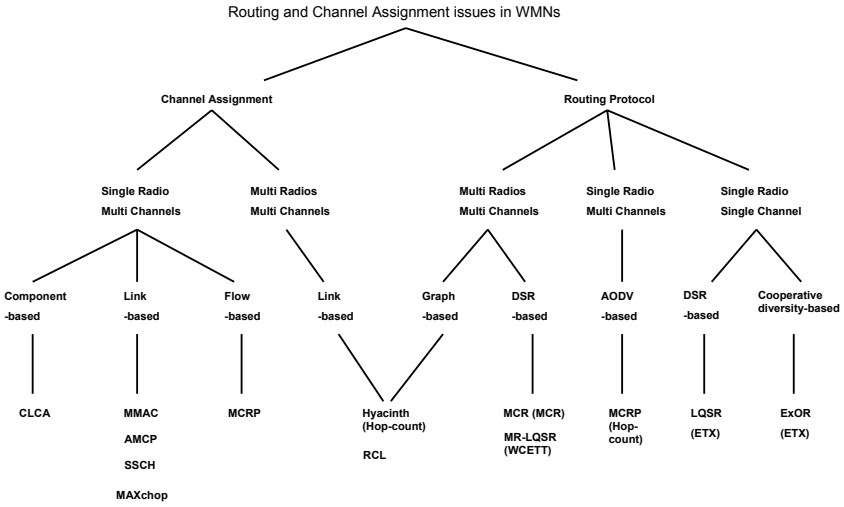



**Fig. 4.5.** A taxonomy of routing and channel assignment protocols for WMNs. Note that the routing metric that a routing protocol uses is given inside parentheses.

where $d_f$ and $d_r$ are the forward and reverse delivery ratio, respectively. With ETX as the route metric, the routing protocol can locate routes with the least expected number of transmissions. Note that the effects of link loss ratios and their asymmetry in the two directions of each link on a path are explicitly considered in the EXT measure. Measurements on wireless testbeds [16, 72] show that, for the source-destination pairs that are with two or more hops, use of ETX as the route metric renders routes with throughput significantly higher than use of the minimum hop count.

*Expected Transmission Time (ETT)*

One major drawback of ETX is that it may not be able to identify high-throughput routes, in the case of multi-radio, multi-rate wireless networks. This is because ETX only considers the packet loss rate on a link but not its bandwidth. ETT has thus been proposed to improve the performance of ETX in multi-radio wireless networks that support different data rates. Specifically, ETT includes the bandwidth of a link in its computation [73], i.e.,

$$ETT = ETX \times \frac{S}{B}$$

where $S$ and $B$ denote the size of the packet and the bandwidth of the link, respectively. ETT considers the actual time incurred in using the channel (excluding the

backoff time incurred in accessing the radio channel). In order to measure the bandwidth $B$ of each link, a node sends two probe packets of different sizes (137 and 1137 bytes) to each of its neighbors every minute. The receiver node measures the difference between the instants of receiving the packets, and forwards the information to the sender. The bandwidth is then estimated by the sender node by dividing the larger packet size by the minimum of 10 consecutive measurements. Measurement results on a testbed show that use of ETT significantly improve the systems performance in a multiple-radio network.

*Weighted Cumulative ETT (WCETT)*

What ETX and ETT have not explicitly considered is the intra-flow interference. WCETT was proposed [73] to reduce the number of nodes on the path of a flow that transmit on the same channel. Specifically, let $X_c$ be defined as the number of times channel $c$ is used along a path. Then WCETT for a path is defined as the weighted sum of the cumulative expected transmission time and the maximal value of $X_c$ among all channels, i.e.,

$$WCETT = (1-\beta)\sum_{i=1}^{n} ETT_i + \beta \max_{1\le c\le C} X_c \tag{4.1}$$

where $\beta$ ($0 \le \beta \le 1$) is a tunable parameter. Reducing the first term of Eq. (4.1) improves the global resource utilization, while reducing the second term of Eq. (4.1) increases the achievable throughput by reducing the intra-flow interference. Moreover, the two terms also represent a trade-off between achieving low delay and high throughput. Reducing the first term reduces the delay, while reducing the second term increases the achievable link throughput. The tunable parameter $\beta$ is used to adjust the relative importance of the two objectives.

*Modified Expected Number of Transmissions (mETX) and Effective Number of Transmissions (ENT)*

Another issue which ETX does not consider is the effect of short-term channel variation, i.e., ETX takes only the average channel behavior into account for the route decision. In order to capture the time-varying property of a wireless channel, the metrics mETX and ENT were proposed in [74] which took into account both the average and the standard deviation of the observed channel loss rates. Specifically, mETX is expressed as

$$mETX = \exp\left(\mu_\Sigma + \frac{1}{2}\sigma_\Sigma^2\right)$$

where $\mu_\Sigma$ and $\sigma_\Sigma^2$ are the average and variability of the channel bit error probability. In some sense, mETX incorporates the impact of physical layer variability in the design of routing metrics. On the other hand, when the problem of maximizing aggregate throughput with the packet loss rate constraint is considered, mETX

may not be sufficient since the links which mETX selects may achieve the maximum link-layer throughput but incur high loss rates at the same time. The ENT metric is devised to meet both objectives. Specifically, ENT is expressed as

$$ENT = \exp\left(\mu_\Sigma + 2\delta\sigma_\Sigma^2\right)$$

where $\delta$ is the strictness of the loss rate requirement. As shown in both experimental and simulation results, mETX and ENT achieve a 50% reduction in the average packet loss rate as compared with ETX. This implies the effect of time-varying channels should be considered in designing a throughput-optimizing route metric.

### 4.4.2 Representative Routing Protocols

*Link Quality Source Routing (LQSR)*

LQSR [16] is a modified version of DSR and aims to select a better route using link-quality metrics in single-radio, single-channel wireless networks. LQSR implements the basic functionalities of DSR including route discovery and route maintenance. In addition, a variety of link quality metrics including ETX, Per-hop Round Trip Time (RTT) [75], Packet Pair [76] and hop count were supported as routing metrics.

LQSR is realized based upon the *Mesh Connectivity Layer (MCL)*, a loadable Microsoft Windows driver. It is located between layer 2 (link layer) and layer 3 (network layer) of the standard ISO/OSI model. To the higher layers, MCL appears to be another Ethernet link although it is a virtual one. To the lower layers, MCL appears to be another protocol running over the physical link. A basic functionality of this protocol is to monitor link quality continuously and change to the path that has the lowest overall cost. Metrics for monitoring the link quality for links actively in use are maintained by using a reactive mechanism.

*Extremely Opportunistic Routing (ExOR)*

ExOR [77] is a routing protocol that heavily leverages MAC attributes for data transfer. It aims to increase the throughput of large unicast transfers in single-radio, single-channel wireless networks. Central to ExOR is the notion of *cooperative diversity routing*. This notion was originally devised to avoid multi-path fading by using broadcasts to send information through multiple relays concurrently. This allows use of links which traditional routing would typically ignore.

ExOR broadcasts each packet and chooses a receiver to forward only after learning the set of nodes that actually receive the packet. ExOR attempts to send the packet as far as possible by selecting as the forwarding node the node that has the least distance to the final destination. In the course of packet forwarding, ExOR uses acknowledgments (ACKs) to ensure that only one node forwards the packet. This broadcast and forwarding approach takes advantage of “lucky” situations in which unanticipated receivers closer to the destination may be able help transport of the packet. In order to realize ExOR, a loss-rate matrix has to be available that contains the probability of successful packet reception between each pair of nodes. Such a

matrix can be built using, for example, a link-state flooding scheme. Every packet is required to include a set of forwarding candidates prioritized by the distance. The forwarding-decision is then based on the set of forwarding candidates found in the header of the received packet, and on the set of received ACKs which are sent following the receipt of the packet. The node which classifies itself as the forwarding node then retransmits the packet, using a new set of forwarding candidates. In addition, ExOR coordinates data sending between nodes with the use of a timed scheduling algorithm that gives preference to higher priority nodes and ensures collisions do not occur.

Experimental results performed on MIT Roofnet show that ExOR improves the throughput by a factor of 2 or 4 over ETX since it uses multiple relay nodes to forward the packet to its destination. It is also shown that the total number of transmissions required to route a packet from a source to its corresponding destination can be improved by 55-65% in comparison to the best predetermined route from the wired model.

*Multi-Channel Routing Protocol (MCRP)*

MCRP [78] is a routing protocol that is specifically designed for multi-channel networks with single-radio nodes and exploits a channel switching technique. MCRP assigns channels to data flows rather than assigning channels to nodes. Thus, all nodes on the path on which a data flow traverses are assigned to a common channel. This approach is well-suited for on-demand routing where channels are assigned in conjunction with the route discovery procedure. The advantage of this approach is that once the route is established, nodes do not need to switch channels for the duration of the flow. Moreover, because this approach attempts to allocate different channels to different flows, it allows simultaneous transmissions and improves network capacity.

In the route discovery phase, a node with packets to send broadcasts a Route Request (RREQ) packet on each channel in a round robin manner. A RREQ packet contains the channel table and the flow table to be propagated to the destination. The channel table contains the number of times a channel has been consecutively used on a single flow path, and the flow table contains the number of times simultaneous flows have been carried out on a single channel. These tables are used by the destination node to select a feasible and load balancing route. Upon receipt of a RREQ packet, a node also rebroadcasts the RREQ (unless it itself is the destination). Moreover, the node also creates a reverse path to the source and maintains the information of the channel on which the RREQ arrives. Upon receipt of one or more RREQ packets, the destination prepares a Route Reply (RREP) packet (that contains the selected channel) and unicasts it on the selected path. All nodes that have forwarding the corresponding RREQ packet change their operating channels to the channel selected by the destination.

*Multi-Radio Link Quality Source Routing (MR-LQSR)*

MR-LQSR is essentially the LQSR protocol with the use of the WCETT metric [73]. Similar to LQSR, MR-LQSR also operates in conjunction with the Mesh Connectivity Layer (MCL). It has three main objectives: (i) the loss rate and the bandwidth of

a link should be taken into account for selecting a path; (ii) the path metric should be increasing; and (iii) the path metric should reflect the throughput degradation due to the interference caused by simultaneous transmissions. Towards these objectives, WCETT is considered as a path metric to account for the interference among links on the same channel.

To incorporate WCETT into LQSR, the information including the channel assigned on a link, its bandwidth and loss rate is propagated to all nodes in the network, in the form of DSR control packets. To calculate WCETT, the ETT on each link is first computed using the ETX, the bandwidth and the packet loss. The ETT metric is then used to compute the WCETT. Finally, the WCETT is applied to the link cache scheme of the DSR protocol. In native DSR, since the default cost of each link is set to one, the Dijkstra algorithm, when executed over the link cache by a source node, always gives the shortest path with the minimum hops. On the other hand, when the WCETT is used as the link cost, it produces the minimal cost path in terms of link bandwidth and loss rate.

*Multi-Channel Routing (MCR)*

MCR [79] is an on-demand, multi-channel routing protocol for WMNs with multi-radio nodes. In order to fully exploit the available channels with a limited number of radios on each node, the protocol uses a switching mechanism to change channels assigned to a radio interface whenever necessary. In particular, two types of interfaces are assumed: fixed and switchable. $K$ interfaces out of a total $M$ interfaces are fixed interfaces and are designated to some $K$ channels. The remaining interfaces are dynamic interfaces and dynamically assigned to any of the remaining channels. Multiple queues are maintained for all switchable interfaces.

Each node maintains a neighbor table and a channel usage list. The neighbor table contains the fixed channels used by the node's neighbors. The channel usage list contains the count of two-hop neighborhoods that are using a channel as their fixed channel. Each node periodically transmits a HELLO packet on all channels, including its fixed channel number and neighbor table. A node receiving the HELLO packet then updates its neighbor table and channel usage list. The table and list information are used for the switching mechanism to make a decision of which channel is assigned to what interface in the link layer. Furthermore, the switching mechanism helps MCR for selecting routes over multiple channels.

The route used in MCR is a weighted sum of two elements. The first element accounts for the resources consumed along the path and is obtained by summing ETT values along the path. Note that because the switching cost is (implicitly) part of the ETT of each link, the first term contains the switching cost. The second term accounts for the channel diversity cost, and is calculated by finding the maximum ETT cost on all channels. Accordingly, a route with a larger number of distinct channels may have a lower diversity cost. Different from WCETT which is designed for the case in which the number of interfaces per node is equal to the number of channels, the MCR metric is applicable to a more general case where the number of available interfaces may be smaller than the number of available channels, and interface switching is needed.

The route discovery phase of MCR is similar to that of DSR. In addition, each RREQ also contains the channel number and the switching cost. Thus, when the RREQ is received by destination, the diversity cost (i.e., the number of channels in the RREQ) and the switching cost (i.e., the sum of all link switching costs) are calculated. Based on these costs, the destination selects the optimal path available between the source and the destination.

*Joint Routing and Channel Allocation*

Alicherry *et al.* [80] proposed a joint routing, channel assignment and link scheduling (RCL) algorithm that attempts to maximize throughput in a multi-channel and multi-radio network. The work is performed under the premise that topology change in WMNs is infrequent and the variability of aggregate traffic demand from each mesh router (client traffic aggregation point) is small. These characteristics allow optimization to be made periodically by the system management software based on traffic demand estimates. Under the assumption that the network is restricted to be a superset of a disk graph (i.e., the interference range is assumed to be a fixed multiple of the communication range), the authors mathematically formulated the joint channel assignment and routing problem taking into account interference constraints, the number of channels in the network, and the number of radios available at each mesh router. They then solved the problem with the use of the LP relaxation technique. This was then followed by (i) several adjustment steps to obtain a valid channel assignment and a link scheduling policy that eliminates interference; and (ii) a post processing phase and a flow scaling round to make the assignment interference-free.

*Hyacinth Network Architecture that Supports Channel Assignment and Routing*

Raniwala and Chiueh [81] proposed a network architecture, called *Hyacinth*, for wireless mesh networks with multi-channels and multi-radios. This architecture supports a fully distributed channel assignment algorithm and a spanning-tree based routing algorithm. The mesh routers having access to the wired network are considered as the root nodes of the spanning tree. Based on the spanning tree, routing is performed to balance traffic load over the network as well as to repair route failures. The channel assignment algorithm operates in two phases: *neighbor interface binding* and *interface-channel assignment*. In the first phase each node classifies its interfaces into the set of network interface cards (NICs) for its parent node termed as UP-NICs and the set of NICs for its children nodes termed as DOWN-NICs. Each node can assign and change the channel on its DOWN-NICs only. The purpose of this phase is to bound the impact of change in channel assignment since the change may cause a series of channel re-assignment across the network. In the second phase each node periodically exchanges messages that contain the channel usage status with its neighbors in the interference range. Based on the status of channels used in the neighborhood, a node then determines a set of channels that are least-used in its vicinity. The advantage of this channel assignment scheme is that it achieves a tree architecture where links close to the root of the spanning tree are given higher bandwidth.

The channel assignment is further combined with the routing process. Each node which has routing information to the root advertises this information to one-hop neighbors containing the cost metrics such as the hop-count and the residual uplink capacity. Based on the cost, each node which receives the advertisement makes a decision with regard to joining the advertising node. If the node decides to join, it sends an acceptance message to the advertising node and a departing message to its parent node with which it is now associated. The joining process for new nodes to the network is initiated by broadcasting HELLO packets to the neighboring nodes.

## 4.5 Open Research Issues

In spite of the bulk of research in the literature, there are still open research issues that should be addressed in order to build high-performance and robust WMNs. In this section, we outline these open research issues.

*Topology Control Under the Physical SINR Model*

As mentioned in Section 4.2.1, most of the studies on topology control are inherently based on the graph model that characterizes graph-theoretic properties of wireless networks, while ignoring important physical aspects of communications. Recently, Moscibroda et al. [30] studied the problem of topology control under an information-theoretic SINR model. They derived the time complexity of a scheduling algorithm that assigns transmit power levels to all the nodes and schedules all links of an arbitrary network topology. They proved that if the signals are transmitted with correctly assigned transmission power levels, the number of time slots required to successfully schedule all links is proportional to the squared logarithm of the network size. They also devised a centralized algorithm for approaching the theoretical upper bound. In spite of its theoretical importance, the centralized scheduling algorithm cannot, however, be practically implemented. Devising *localized* topology control algorithms under the physical SINR model remains as a research challenge.

*Channel Assignment and Routing in Multi-radio, Multi-channel Environments*

A traditional channel assignment problem is what channel should be assigned to a transmission pair in order to enable transmission, mitigate inter-/intra-interference, and improve network capacity. This problem is augmented with another dimension in multi-radio and multi-channel environments: what channel should be associated with each of the radio interfaces a node possesses? Although there have been some preliminary work [79, 82], a rigorous treatment of this problem has been lacking. This problem is further complicated, when it is considered in conjunction with routing. Several research efforts [79–81] have been made to address the joint problem of channel assignment and routing, and various heuristics (although with insightful theoretical base) have been proposed under certain (perhaps unrealistic) interference models. The challenge, however, remains to consider the problem in an analytic framework under a realistic interference model (in which cumulative interference due to concurrent transmissions is faithfully characterized).

*Tuning All the PHY/MAC Control Knobs for Spatial Reuse*

As mentioned in Section 4.2, there are several PHY/MAC attributes that can be used to improve spatial reuse, mitigate interference and and maximize network capacity: (i) the *transmit power* each node uses for communications, (ii) the *carrier sense threshold* each node uses to determine if the shared medium is idle, (iii) the *channel* on which the node transmits, and (iv) the *time intervals* in which each node gain access to the channel. On top of all these, routing also plays an important role in mitigating interference and improving end-to-end throughput. Most existing work has only focused on tuning one or two attributes, in spite of the fact that these attributes actually interwined with each other. The challenge remains to establish an optimization framework of maximizing the network capacity by adjusting PHY/MAC parameters in all possible dimensions in the design space.

*Routing Metrics that Leverage PHY/MAC Attributes*

As discussed in Section 4.4.1, several routing metrics have been proposed based on the link transmission time (estimated by probe packets). There are, however, a much richer set of PHY/MAC attributes that can be leveraged for cross-layer design and implementation (Section 4.3). Incorporating some of these PHY/MAC attributes in the calculation of routing metrics may render better, higher-throughput routes and further improve the overall network performance.

*Overheads Incurred in Cross-Layer Design and Optimization*

Most of the theoretical results that demonstrate the advantage of cross-layer design and optimization in WMNs do not adequately consider the computing and communications overhead thus incurred, i.e., the overhead incurred in collecting information needed for inferring the interference, calculating the route metrics, switching the channels, or scheduling frame transmission. It is thus not clear whether or not the performance gain in engaging multiple protocol entities in the protocol stack or across the network outweighs the overhead thus incurred. An in-depth empirical study on a large WMN is needed to better quantify the overhead.

*Considering Mesh Client Characteristics in WMNs*

In WMNs, there are roughly two entities: mesh routers and mesh clients. The former is usually stationary and not energy-constrained, while the latter is battery-powered and may move arbitrarily. Most of the existing studies have focused on MAC and routing on mesh routers, without considering the characteristics of mesh clients. Incorporating the end-to-end performance requirements and constraints of mesh clients into WMN design will be an interesting and challenging research issue.

## Conclusion

By virtue of their robustness, cost-effectiveness, self-organizing and self-configuring nature, WMNs have emerged as a new network paradigm for a wide range of applications, such as public safety and emergency response communications, intelligent transportation systems, and community networks. One fundamental problem of WMNs with a limited number of radio interfaces and orthogonal channels is that the performance degrades significantly as the network size grows. This results from increased interference between nodes and diminished spatial reuse over the network.

In this chapter, we have addressed several research issues pertinent to the performance and capacity optimization issues in WMNs. We have provided a taxonomy of recent advances in the literature with respect to radio resource management (adjusting transmission rate, power, carrier sense threshold and assigning channels) and routing. We have also addressed important issues regarding design and implementation of device driver support that facilitates cross layer design and optimization. Finally, we have outlined several research avenues in which future research can pursue.

## References

1. Microsoft Networking Research Group, "Self-organizing neighborhood wireless mesh networks," http://research.microsoft.com/mesh/.
2. Nortel Networks, "The business case for wireless mesh networks," http://www.nortelnetworks.com/corporate /events/2003d/ wmn_eseminar/col lateral/wmn_eseminar.pdf, December 2003.
3. Bay Area Wireless Users Group, http://www.bawug.org/.
4. CUWiN, http://www.cuwireless.net/.
5. SFLan, http://www.sflan.org.
6. Seattle Wireless, http://www.seattlewireless.net/.
7. Southampton Open Wireless Network, http://www.sown.org.uk/.
8. Wireless Leiden, http://www.wirelessleiden.nl/.
9. MIT Roofnet, http://pdos.csail.mit.edu/roofnet/doku.php.
10. Technology for All Wireless, http://tfa.rice.edu/.
11. D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris, "Link-level measurements from an 802.11b mesh network," in *Proc. of ACM SIGCOMM*, September 2004.
12. J. Bicket, D. Aguayo, S. Biswas, and R. Morris, "Architecture and evaluation of an 802.11b mesh network," in *Proc. of ACM MobiCom*, September 2005.
13. Z. Fu, H. Luo, P. Zerfos, S. Lu, L. Zhang, and M. Gerla, "The impact of multihop wireless channel on TCP performance," *IEEE Trans. on Mobile Computing*, vol. 4, no. 2, pp. 209–221, March/April 2005.
14. D. Berger, Z. Ye, P. Sinha, S. Krishnamurthy, M. Faloutsos, and S. K. Tripathi, "TCP-friendly medium access control for ad-hoc wireless networks: Alleviating self-contention," in *Proc. of IEEE MASS*, October 2004.
15. K. Sanzgiri, I. D. Chakeres, and E. M. Belding-Royer, "Determining intra-flow contention along multihop paths in wireless networks," in *Proc. of Broadnets Wireless Networking Symposium*, October 2004.

16. R. Draves, J. Padhye, and B. Zill, "Comparison of routing metrics for static multi-hop wireless networks." in *Proc. of ACM SIGCOMM*, 2004.
17. A. Akella, G. Judd, P. Steenkiste, and S. Seshan, "Self management in chaotic wireless deployments," in *Proc. of ACM MobiCom*, 2005.
18. A. Miu, H. Balakrishnan, and C. E. Koksal, "Achieving loss resiliency through multi-radio diversity in wireless networks," in *Proc. of ACM MobiCom*, 2005.
19. L. Li, J. Halpern, V. Bahl, Y. Wang, and R. Wattenhofer, "Analysis of a cone-based distributed topology control algorithm for wireless multi-hop networks," in *Proc. of ACM Symposium on Principle of Distributed Computing (PODC)*, 2001, pp. 264–273.
20. N. Li and J. C. Hou, "FLSS: A fault-tolerant topology control algorithm for wireless networks," in *Proc. of ACM MobiCom*, 2004.
21. ——, "Topology control in heterogeneous wireless networks: Problems and solutions," in *Proc. of IEEE INFOCOM*, March 2004.
22. ——, "Localized topology control algorithms for heterogeneous wireless networks," *IEEE/ACM Trans. on Networking*, vol. 13, no. 6, pp. 1313–1324, December 2005.
23. ——, "Localized fault-tolerant topology control in wireless ad hoc networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 4, pp. 307–320, April 2006.
24. N. Li, J. C. Hou, and L. Sha, "Design and analysis of a MST-based distributed topology control algorithm for wireless ad-hoc networks," *IEEE Trans. on Wireless Communications*, vol. 4, no. 3, pp. 1195–1207, 2005.
25. R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit power adjustment," in *Proc. of IEEE INFOCOM*, March 26–30, 2000.
26. V. Rodoplu and T. Meng, "Minimum energy mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1333–1344, August 1999.
27. R. Wattenhofer, L. Li, P. Bahl, and Y.-M. Wang, "Distributed topology control for power efficient operation in multihop wireless ad hoc networks," in *Proc. of IEEE INFOCOM*, April 22–26, 2001.
28. R. Prim, "Shortest connection networks and some generalizations," *The Bell System Technical Journal*, vol. 36, pp. 1389–1401, 1957.
29. M. Burkhart, P. Rickenbach, R. Wattenhofer, and A. Zollinger, "Does topology control reduce interference?" in *Proc. of ACM MobiHoc*, May 2004.
30. T. Moscibroda, R. Wattenhofer, and A. Zollinger, "Topology control meets SINR: The scheduling complexity of arbitrary topologies," in *Proc. of ACM MobiHoc*, June 2006.
31. Y. Gao, J. C. Hou, and H. Nguyen, "Physical-model-based topology control for maintaining network connectivity and maximizing network capacity," in *Proc. of IEEE Int'l Conf. on Network Protocols*, October 2007.
32. H. Lim, C. Lim, and J. C. Hou, "A coordinate-based approach for exploiting temporal-spatial diversity in wireless mesh networks," in *Proc. of ACM MobiCom*, September 24–29, 2006.
33. J. Monks, V. Bharghavan, and W.-M. Hwu, "A power controlled multiple access protocol for wireless packet networks," in *Proc. of IEEE INFOCOM*, April 22–26, 2001.
34. A. Muqattash and M. Krunz, "Power controlled dual channel (PCDC) medium access protocol for wireless ad hoc networks," in *Proc. of IEEE INFOCOM*, March 30–April 3, 2003.
35. ——, "A single-channel solution for transmission power control in wireless ad hoc networks," in *Proc. of ACM MobiHoc*, May 24–26, 2004.
36. S. Narayanaswamy, V. Kawadia, R. Sreenivas, and P. R. Kumar, "Power control in ad-hoc networks: Theory, architecture, algorithm and implementation of the COMPOW protocol," in *European Wireless Conference*, 2002.

37. T. Nadeem, L. Ji, A. Agrawala, and J. Agre, "Location enhancement to IEEE 802.11 DCF," in *Proc. of IEEE INFOCOM*, March 13–17, 2005.
38. A. Vasan, R. Ramjee, and T. Woo, "ECHOS: Enhanced capacity 802.11 hotspots," in *Proc. of IEEE INFOCOM*, March 13–17 2005.
39. J. Zhu, X. Guo, L. Yang, and W. S. Conner, "Leveraging spatial reuse in 802.11 mesh networks with enhanced physical carrier sensing," in *Proc. of IEEE ICC*, June 20–24, 2004.
40. J. Zhu, X. Guo, L. Yang, W. S. Conner, S. Roy, and M. M. Hazra, "Adapting physical carrier sensing to maximize spatial reuse in 802.11 mesh networks," *Wiley Wireless Communications and Mobile Computing*, vol. 4, pp. 933–946, December 2004.
41. X. Yang and N. H. Vaidya, "On the physical carrier sense in wireless ad hoc networks," in *Proc. of IEEE INFOCOM*, March 13–17, 2005.
42. Z. Zeng, Y. Yang, and J. C. Hou, "How physical carrier sense affect protocol capacity: modeling and analysis," Dept. of Computer Science, University of Illinois at Urbana Champaign, Tech. Rep., April 2007.
43. F. Calì, M. Conti, and E. Gregori, "IEEE 802.11: Design and performance evaluation of an adaptive backoff mechanism," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 9, September 2000.
44. H. Zhai and Y. Fang, "Physical carrier sensing and spatial reuse in multirate and multihop wireless ad hoc networks," in *Proc. of IEEE INFOCOM*, April 23–29, 2006.
45. J. A. Fuemmeler, N. H. Vaidya, and V. V. Veeravalli, "Selecting transmit powers and carrier sense thresholds for CSMA protocols," Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Tech. Rep., October 2004.
46. T.-S. Kim, H. Lim, and J. C. Hou, "Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks," in *Proc. of ACM MobiCom*, September 24–29, 2006.
47. Y. Yang, J. C. Hou, and L.-C. Kung, "Modeling of physical carrier sense in multi-hop wireless networks and its use in joint power control and carrier sense adjustment," in *Proc. of IEEE INFOCOM Miniconferences*, May 6–12, 2007.
48. G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, March 2000.
49. A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed point analysis of single cell IEEE 802.11 WLANs," in *Proc. of IEEE INFOCOM*, April 2005.
50. IEEE Computer Society, "Part 11: Wireless LAN Medium Access (MAC) and Physical Layer (PHY) Specifications, Amendment 8: Medium Access Cntorl (MAC) Quality of Service Enhancement," *IEEE Standard 802.11e*, 2005.
51. J. So and N. H. Vaidya, "Multi-channel MAC for ad hoc networks: Handling multi-channel hidden terminals using a single transceiver," in *Proc. of ACM MobiHoc*, 2004.
52. J. Shi, T. Salonidis, and E. Knightly, "Starvation mitigation through multi-channel coordination in CSMA multi-hop wireless networks," in *Proc. of ACM MobiHoc*, May 2006.
53. A. Mishra, V. Shrivastava, D. Agarwal, S. Banerjee, and S. Ganguly, "Distributed channel management in uncoordinated wireless environments," in *Proc. of ACM MobiCom*, September 2006.
54. R. Vedantham, S. Kakumanu, S. Lakshmanan, and R. Sivakumar, "Component based channel assignment in single radio, multichannel ad hoc networks," in *Proc. of ACM MobiCom*, September 2006.
55. P. Bahl, R. Chandra, and J. Dunagan, "SSCH: Slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks," in *Proc. of ACM MobiCom*, 2004.

56. J. Bicket, "Bit-rate selection in wireless networks," Master's thesis, MIT, 2005.
57. I. Haratcherev, K. Langendoen, R. Lagendijk, and H. Sips, "Hybrid rate control for IEEE 802.11," in *Proc. of ACM MobiWac*, 2004.
58. ——, "Fast 802.11 link adaptation for real-time video streaming by cross-layer signaling," in *Proc. of ISCAS*, 2005.
59. G. Holland, N. H. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Proc. of ACM MobiCom*, 2001.
60. A. Kamerman and L. Monteban, "WaveLAN II: A high-performance wireless LAN for the unlicensed band," *Bell Labs Technical Journal*, 1997.
61. J. Kim, S. Kim, S. Choi, and D. Qiao, "CARA: Collision-aware rate adaptation for IEEE 802.11 WLANs," in *Proc. of IEEE INFOCOM*, 2006.
62. M. Lacage, M. H. Manshaei, and T. Turletti, "IEEE 802.11 rate adaptation: A practical approach," in *Proc. of ACM MSWiM*, 2004.
63. D. Qiao and S. Choi, "Fast-responsive link adaptation for IEEE 802.11 WLANs," in *Proc. of IEEE ICC*, 2005.
64. D. Qiao and K. Shin, "Achieving efficient channel utilization and weighted fairness for data communications in IEEE 802.11 WLAN under the DCF," in *Proc. of International Workshop on Quality of Service 2002*, May 2002.
65. B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," in *Proc. of ACM MobiCom*, 2002.
66. S. H. Y. Wong, H. Yang, S. Lu, and V. Bharghavan, "Robust rate adaptation for 802.11 wireless networks," in *Proc. of ACM MobiCom*, September 24–29, 2006.
67. L.-C. Kung, J. Choi, J. C. Hou, Y. Gao, I.-H. Hou, R. Lam, and Y. Yang, "Toward building an open wireless mesh network for cross layer design and optimization," Dept. of Computer Science, University of Illinois at Urbana-Champaign, Tech. Rep., May 2007.
68. MADWIFI, http://sourceforge.net/projects/madwifi.
69. V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *Proc. of IEEE INFOCOM*, 1997.
70. C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proc. of ACM SIGCOMM*, 1994.
71. C. E. Perkins and E. M. Royer, "Ad-hoc on demand distance vector routing," in *Proc. of IEEE Workshop on Mobile Computing Systems and Applications*, 1999.
72. D. S. J. De Couto, D. Aguayo, J. C. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proc. of ACM MobiCom*, 2003.
73. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. of ACM MobiCom*, 2004.
74. C. Koksal and H. Balakrishnan, "Quality-aware routing metrics for time-varying wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 1984–1994, Nov. 2006.
75. A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "A multi-radio unification protocol for IEEE 802.11 wireless networks," in *Proc. of IEEE BroadNets*, July 2004.
76. S. Keshav, "A control-theoretic approach to flow control," *Proc. of the conference on Communications architecture and protocols*, pp. 3–15, 1993.
77. S. Biswas and R. Morris, "ExOR: Opportunistic routing in multi-hop wireless networks," in *Proc. of ACM SIGCOMM*, September 2005.
78. J. So and N. H. Vaidya, "Routing protocol for utilizing multiple channels in multi-hop wireless networks with a single transceiver," Dept. of Computer Science and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Tech. Rep., 2004.

79. P. Kyasanur and N. H. Vaidya, "Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks," University of Illinois at Urbana-Champaign, Tech. Rep., May 2005.
80. M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. of ACM MobiCom 2005*, September 2005.
81. A. Raniwala and P. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. of IEEE INFOCOM*, 2005.
82. M. Benveniste, "The CCC mesh MAC protocol," IEEE 802.11-05/0610r1, 2005.

# 5

# Channel Assignment Strategies for Wireless Mesh Networks

M. Conti[1], S. K. Das[2], L. Lenzini[3], and H. Skalli[4]

[1] Italian National Research Council (CNR), Italy
Marco.Conti@iit.cnr.it
[2] The University of Texas at Arlington, USA
das@cse.uta.edu
[3] University of Pisa, Italy
l.lenzini@iet.unipi.it
[4] IMT Lucca Institute for High Studies, Italy habiba.skalli@imtlucca.it

## 5.1 Introduction

A wireless mesh network (WMN), as illustrated in Fig. 5.1, consists of mesh routers and mesh clients. The mesh routers are generally stationary nodes and form a multi-hop wireless backbone (referred to as the *backhaul tier*) between the mesh clients and the Internet gateways (a gateway is the node directly connected to the wired network). Each mesh router operates not only as a host but also as a router, forwarding packets on behalf of other nodes that may not be within direct wireless transmission range of their destinations. On the other hand, mesh clients form the *client tier*. They are either stationary or mobile, and can form a client mesh network with each other and with mesh routers. The gateway and bridge functionalities in mesh routers enable the integration of WMNs with various existing wireless networks such as wireless sensor, cellular, wireless-fidelity (Wi-Fi), and worldwide inter-operability for microwave access (WiMAX).

WMNs have emerged as a highly flexible, reliable and cost efficient solution for wirelessly covering large areas and for providing low-cost Internet access through multi-hop communications. It is anticipated that they will not only resolve the limitations of wireless ad hoc networks, local area networks (WLANs), personal area networks (WPANs), and metropolitan area networks (WMANs) but also significantly improve such networks' performance. Several emerging and commercially interesting applications for commodity networks based on the WMN architecture have also been deployed, see [3]. They include community and neighborhood networks, broadband home networking, enterprise networking, building automation, intelligent transportation systems, public safety networks, etc. Perhaps among the earliest and the most important of these are community and neighborhood networks. The networking solution based on WMNs mitigates many of the disadvantages of the conventional WLAN architecture based on a digital subscriber line (DSL) where the last

hop is wireless. For example, within the WLAN scenario, even if information has to be shared within a community or neighborhood, all traffic must flow through the Internet. Moreover, only a single path may be available for one house to access the Internet. Additionally, wireless services must be set up individually at every home. As a result, network service costs may increase [1]. Deployment of a WMN is a robust and inexpensive alternative; the wireless backbone has the ability to support both internal (among mesh routers) and external (to the Internet) traffic. It also guarantees the existence of multiple paths and makes it possible to cover larger areas with lower costs.

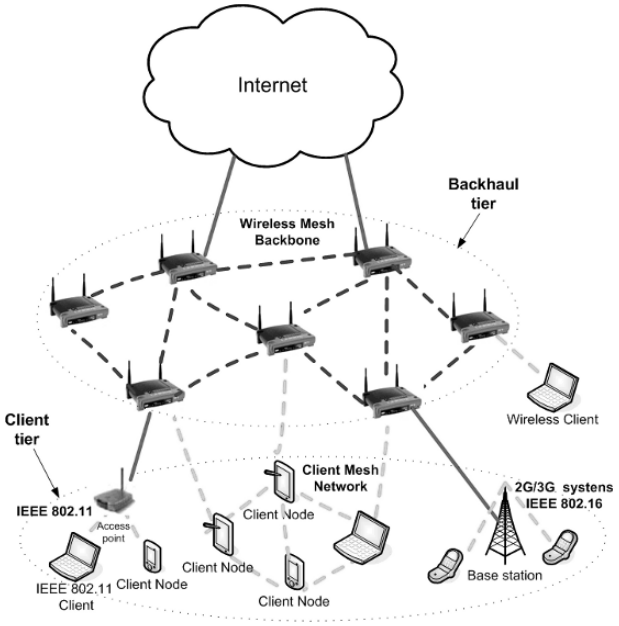



**Fig. 5.1.** Wireless mesh network architecture.

However, the major technical challenges (i.e. capacity, scalability) of building a large-scale high-performance multi-hop wireless mesh networks have not been solved yet. Wireless mesh networks [19], which use off-the-shelf 802.11 based network cards[5], are typically configured to operate on a single channel (part of the frequency spectrum with a specified bandwidth) using a single radio. This configuration adversely affects the capacity of the mesh due to interference from adjacent nodes in the network (i.e. all neighboring nodes will compete on the same channel).

There are several on-going research efforts to improve the capacity of wireless mesh networks by exploiting such alternative approaches as multiple radio interfaces

[5]Throughout this chapter, the terms interface and network interface card (NIC) will have equivalent meaning to 'radio'.

[6], directional antennas [4], multiple-input multiple-output (MIMO) techniques [1], and modified medium access control (MAC) protocols adapted for WMNs [16]. By using directional transmission, the interference between network nodes can be mitigated, and thus the network capacity can be improved [27]. Directional antennas can also improve energy efficiency [34]. However, they bring challenges to the MAC protocol design [15, 23]. The MIMO technique consists of using multiple antennas in both the transmitter and the receiver. MIMO deploys simultaneous transmissions and transmit/receive diversity (receive diversity is when the same information is received by different antennas; transmit diversity is when the same information is sent from multiple transmit antennas). Thus, MIMO can potentially increase the system's capacity [20]; however, in this case also an efficient MAC protocol exploiting MIMO characteristics is needed to achieve significant throughput improvement. As far as the MAC protocols are concerned, scalability is still a very challenging issue for designing an efficient MAC protocol for WMNs. Most of the existing MAC protocols partially solve the problem, but raise other problems such as throughput, capacity or fairness [3]. Moreover, a MAC protocol for WMNs must consider both scalability and heterogeneity between different network nodes (i.e. mesh routers, mesh clients).

Equipping each node with multiple radios is emerging as a promising approach for improving the capacity of WMNs. First, the IEEE 802.11b/g [11] and IEEE 802.11a [10] standards provide 3 and 12 non-overlapping channels, respectively, which can be used simultaneously by a mesh router for transmission and reception within a neighborhood by tuning non-overlapping channels to different radios. This then leads to efficient spectrum utilization and increases the actual bandwidth available to the network. Secondly, the availability of cheap, off-the-shelf commodity hardware also makes multi-radio solutions economically attractive. Finally, the spatio-temporal diversity of radios operating on different frequencies with different sensing-to-hearing ranges, bandwidth, and fading characteristics can be leveraged to improve the overall capacity of the network.

In a realistic WMN, the total number of radios is much higher than the number of available channels. Thus, many links between the mesh routers will be operating on the same set of channels. At the same time, interference among transmissions on these channels can dramatically decrease their utilization (e.g. due to contention among the nodes, as in the IEEE 802.11 protocol). Therefore, as with cellular networks, the key factor for minimizing the effect of interference is the efficient reuse of the scarce radio spectrum. Therefore, a key issue in a multi-radio, multi-channel WMN architecture is the channel assignment problem which involves assigning (binding) each radio to a channel in such a way that efficient utilization of available channels can be achieved. Specifically, the channel assignment problem in multi-hop communication is targeted at minimizing interference on any given channel. In addition, another fundamental goal of WMN channel assignment is to guarantee an adequate level of connectivity among the mesh nodes. In other words, the assignment of channels to radios should ensure that multiple paths are available among mesh routers. This is a major characteristic and requirement for the robustness and reliability of the WMN backhaul tier.

A WMN node needs to share a common channel with each of its neighbors in the communication range, requiring it to set up a virtual link[6]. Moreover, to reduce network interference, a node should minimize the number of neighbors that it shares a common channel with. Therefore, there exists a trade-off between maximizing connectivity and minimizing interference. This trade-off is illustrated by the example in Fig. 5.2. Fig. 5.2(a) shows the connectivity of the network when a single channel is operating on a single radio. In this scenario, a link is placed between two nodes if they are within their respective transmission ranges.

This is the maximum achievable network connectivity since a single common channel is shared between all the nodes. Now, let us focus on the multi-channel multi-radio scenario represented in Figs. 5.2(b) and (c). There are four non-overlapping channels available for communication, given that every node is equipped with two radios. Let us illustrate a case where network connectivity is maximized (same as single radio single channel connectivity), and another case where the interference is minimized (with the efficient use of the available channels). We will also explain how one affects the other. In Fig. 5.2(b), the assignment of channels to the radios results in maximum network connectivity. However, this cannot be achieved unless at most three of the four available channels are assigned and three of the links are assigned the same channel (i.e. channel 2). For instance, there is a direct communication link between every pair of neighbors. However, not all the links can be active simultaneously because of possible interference. On the other hand, Fig. 5.2(c) shows how interference could be completely eliminated and all links can be simultaneously active. The compromise here is that there is no common channel between neighbors, ***b*** and ***d***.

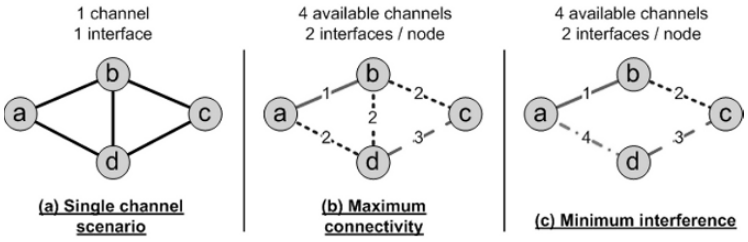



**Fig. 5.2.** Trade-off between connectivity and interference.

The above example clearly illustrates that the goal of channel assignment is to achieve a balance between (i) minimizing interference (on any given channel), and (ii) maximizing connectivity. In this sense, channel assignment in a multi-hop wireless network can be viewed as a topology control problem [21] (similar to transmission power control, for example). Unlike a wired network, links in a wireless network

[6]A virtual link between two nodes is defined as a possible direct communication link between them.

are flexible and can be tuned or configured. The tunable parameters in a wireless environment include channel frequency, transmission power, bit rate, and directional transmission (using directional antennas) [21]. In general, topology control exploits these parameters in order to obtain a desired topology of the network. This can be one of the roles of channel assignment in WMNs in addition to maximizing connectivity and minimizing interference.

In this chapter, we will address the channel assignment problem in multi-radio WMNs which entails assigning a channel to each radio in order to ensure the efficient utilization of the available channels. The rest of this chapter is organized as follows. Section 5.2 discusses the differences and challenges of channel assignment in wireless mesh networks compared to cellular networks. In Section 5.3, we give the necessary background; then in Section 5.4, we highlight the associated constraints and challenges. In Section 5.5, a taxonomy is presented to categorize various channel assignment schemes proposed in the literature, followed by various details on how they work along with examples. Finally, Section 5.6 provides a comparison of the channel assignment algorithms.

## 5.2 Channel Assignment in Cellular Networks vs. WMNs

The channel assignment (CA) problem has been extensively studied in the context of wireless cellular networks [14]. The basic concept used is to divide the radio spectrum into a set of non-interfering disjoint radio channels. These channels can then be used simultaneously whilst maintaining an acceptable adjacent channel separation.

Various techniques are used to divide the radio spectrum, such as frequency division (FD), time division (TD) or code division (CD). In FD, the spectrum is divided into disjoint frequency bands. While in TD, channel separation is achieved by dividing the channel usage into time slots. A combination of FD and TD can also be used to divide each frequency band into time slots.

Let $S_i(k)$ be the set $\boldsymbol{i}$ of wireless terminals, which communicate with the base station using the same channel $\boldsymbol{k}$. Because of the scarcity of the radio spectrum, there is a limited number of channels; thus the same channel $\boldsymbol{k}$ can be reused simultaneously by another set $\boldsymbol{j}$ if the members of sets $\boldsymbol{i}$ and $\boldsymbol{j}$ are spaced enough. These sets, which use the same channel, are called co-channels. The concept of channel reuse is illustrated in Fig. 5.3, where there are seven orthogonal channels available (labeled A to G). Each channel is used for communication inside one cell and is reused simultaneously by another cell that is far enough.

The minimum distance at which co-channels can be reused with acceptable interference is called the co-channel reuse distance. This is possible because due to path loss, the average power received from a transmitter at distance $\boldsymbol{d}$ is proportional to $P_T d^{-\alpha}$, where $\alpha$ is in the range 3-5 depending on the physical environment and $P_T$ is the average transmitter power. The co-channel interference caused by frequency reuse is the most restraining factor on the system's capacity. Therefore, the role of a channel assignment scheme is to minimize this interference by adjusting (i) the distance between co-channels and/or (ii) the transmitter power level. These two methods

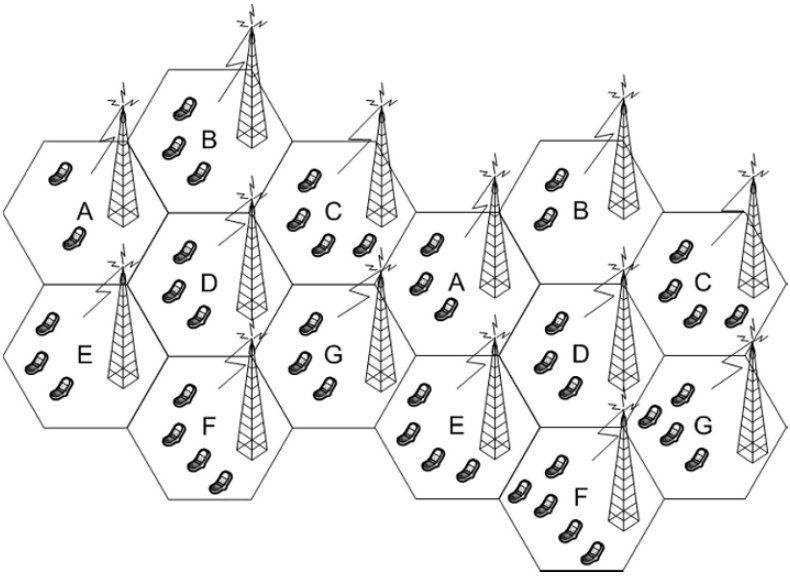



**Fig. 5.3.** The channel reuse concept in cellular networks.

(i and ii) present the underlying concept for channel assignment in cellular systems whose goal is to minimize the carrier-to-interference ratio (CIR) and hence increase radio spectrum reuse efficiency.

In contrast to this, the channel assignment problem in WMNs is different in terms of several aspects. First of all, the architecture of WMNs is different from that of cellular networks. In a WMN, the mesh routers form a multi-hop wireless backbone between mesh clients and the wired network. Whilst in a cellular network, the end-user terminals communicate directly through a single hop with the base-station, and base-station to base-station communication is carried over a separate network which is not the concern of channel assignment.

Secondly, channel assignment in WMNs is mainly aimed at minimizing interference in the wireless backbone. The backhaul is the main focus of research in capacity improvement in WMNs. Channel assignment in cellular networks, on the other hand, is only concerned with minimizing interference on the last hop wireless communication between the base station and the end-user mobile devices and vice versa.

In addition, frequency hopping (FH) is a commonly used technique in cellular networks and consists of rapidly switching frequencies during radio transmission by the base station. FH has many advantages, especially in reducing the effect of noise and interference. This technique could possibly be used in WMNs, however with the current IEEE 802.11 hardware standard, the switching time latency is still extremely high [30] (e.g. in the order of milliseconds). Therefore, such channel switching is difficult, and this makes channel assignment in WMNs more challenging.

## 5.3 Preliminaries

Before we present a taxonomy of the existing channel assignment strategies in WMNs, let us first provide some background concepts and definitions.

### 5.3.1 Connectivity Graph

For modeling purposes, we will consider a WMN with mesh routers[7] distributed on a plane. Each mesh router is equipped with one or multiple radios with omni-directional antennas. We assume that all radios are characterized by an identical transmission range (***R***) and also by the same interference range (***R'***). The *transmission range* is defined as the distance at which a neighbor can receive packet transmissions successfully. When a receiver is within the transmission range of two transmitters that are transmitting simultaneously, the packets are assumed to interfere with each other. This then leads to a collision at the receiver, and thus no packet is received successfully. The *interference range* is defined as the distance at which packet transmission cannot be decoded successfully at the receiver. However, any new transmission from a router within interference range from the receiver interferes with the packet reception. It is generally assumed that the transmission range is smaller than the interference range ($\boldsymbol{R} < \boldsymbol{R'}$) [5].

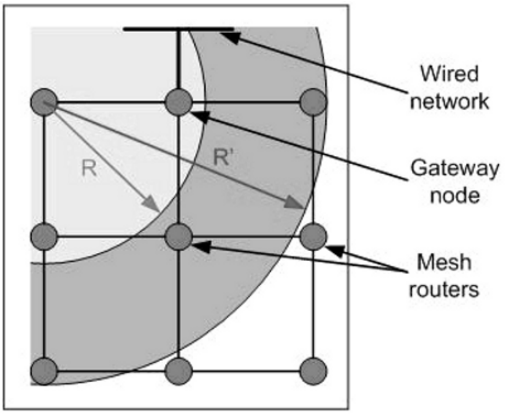



**Fig. 5.4.** An example of a connectivity graph.

According to the above assumptions, connectivity between mesh routers can be modeled using an undirected graph referred to as a *connectivity graph*, ***G***. As illustrated in Fig. 5.4, two nodes in the connectivity graph are linked if they are located within transmission range of each other (see the protocol model, explained in the next

[7] We use the terms mesh router and mesh node interchangeably to refer to the stationary mesh routers that constitute the WMN backbone.

subsection). In general, the network topology (also called logical topology) differs from the connectivity graph, since: a) a link in the connectivity graph may be absent in the network topology graph if the nodes at the end points of this link do not have any radios assigned to a common channel; and b) a link in the connectivity graph may have several corresponding links in the network topology graph if the nodes at the end points have more than one radio each with common channels. Note that the links present in the network topology are referred to as the *logical links*.

### 5.3.2 Conflict Graph

Because of the broadcast nature of the wireless medium, the success of a transmission is greatly influenced by the amount of multiple access interference. This interference can be modeled using a *conflict graph* derived on the basis of a connectivity graph. The concept of a conflict graph is illustrated in Fig. 5.5, where a link between nodes ***x*** and ***y*** in the connectivity graph of Fig. 5.5(a) is represented by a vertex $l_{xy}$ in the conflict graph of Fig. 5.5(b). We use the terms "node" and "link" with reference to the connectivity graph and reserve the terms "vertex" and "edge" for the conflict graph, as in [12]. An edge is placed between two vertices in the conflict graph if the corresponding links in the connectivity graph interfere. The existence and extent of interference between a pair of links are determined by an *interference model*. There are two well-known interference models: (i) the *protocol model*, and (ii) the *physical model*. The protocol model is the simplest and the most commonly used to represent the interference (see Fig. 5.20) whereas the physical model is more complex but offers a more realistic paradigm. Assuming that all nodes in the network have the same interference range, the transmission from ***x*** to ***y*** is successful only if no other node located within distance ***R'*** from ***y*** transmits at the same time as ***x***. Moreover, in the case of IEEE 802.11, if the RTS/CTS (Request to Send/Clear to Send) mode is used, then also no other node within distance ***R'*** from ***x*** should be transmitting at the same time. Therefore, the conflict graph for the protocol model contains an edge between two vertices (i.e. $l_{xy}$, $l_{xz}$) if either ***x*** or ***y*** are located within distance ***R'*** from ***z***.


**Fig. 5.5.** Example illustrating the concept of conflict graph: (a) connectivity graph and corresponding (b) conflict graph.

On the other hand, in the physical interference model, conflicts are not represented as binary. Suppose node ***x*** wants to transmit to node ***y***. The signal strength $SS_{xy}$ of ***x***'s transmission is calculated as received at ***y***. The transmission is successful if $SNR_{xy} \geq SNR_{tresh}$, where $SNR_{xy}$ is the signal to noise ratio at ***y*** of the transmission received from ***x***. The total noise $N_y$ at ***y*** is the total of the ambient noise ($N_a$) and the interference due to other ongoing transmissions in the network. Based on this model, a link $l_{xy}$ exists between ***x*** and ***y*** in the connectivity graph if and only if $SS_{xy}/N_a \geq SNR_{tresh}$ (i.e. SNR exceeds the minimum threshold at least in the presence of ambient noise only). Because conflicts are not binary, interference in the physical model gradually increases as more neighboring nodes transmit and becomes unacceptable when the noise level reaches a threshold. This gradual increase implies that the conflict graph should be a weighted graph, where the weight of a directed edge between two vertices indicates the fraction of the permissible noise at the receiving node. For further details on the physical model see [12].

### 5.3.3 Multi-Radio Conflict Graph

The multi-radio conflict graph (MCG) [26] is an extension of the conflict graph described in the previous subsection. In the MCG, instead of representing the links between mesh routers, vertices represent the links between mesh radios. To create the MCG, each radio in the mesh is represented by a node in a new graph ***G'*** instead of representing routers by nodes as in ***G***.

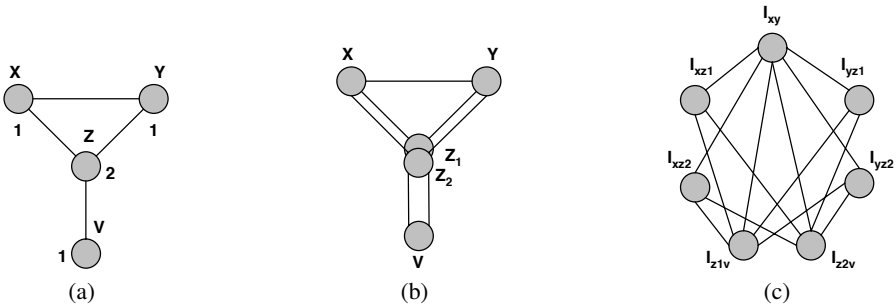



**Fig. 5.6.** An example illustrating multi-radio conflict graph: (a) connectivity graph (**G**), (b) multi-radio connectivity graph (**G'**), and (c) multi-radio conflict graph.

In the above example, let us assume node ***z*** has two radios and the rest of the nodes have one radio as shown in Fig. 5.6(a). Therefore, node ***z*** will be represented by two nodes in ***G'*** as in Fig. 5.6(b), corresponding to its two radios, instead of just one node as in ***G***. Then each link in ***G'*** is represented using a vertex in the MCG. The edges between the vertices in the MCG are created the same way as in the original conflict graph. Two vertices in the MCG have an edge between them if the links

in ***G'*** represented by these two vertices interfere. Fig. 5.6(c) shows the MCG of the wireless mesh network represented in Fig. 5.6(a). In this figure, each vertex is labeled using the radios that make up the vertex. For example, vertex $xz_2$ represents the link between the radio on router ***x*** and the second radio on router ***z***.

## 5.4 Constraints and Challenges in Channel Assignment (CA)

Given the connectivity graph and the interference model, the main challenge for channel assignment is: how to assign a (frequency) channel to each radio in such a way as to minimize interference and maximize connectivity among the nodes. The main constraints [28] that a channel assignment algorithm should satisfy are:

1. The total number of channels is fixed.
2. The number of distinct channels that can be assigned to a mesh router is limited by the number of its radios.
3. Two nodes that share a virtual link expected to carry certain amount of traffic should be bound to a common channel.
4. The sum of the expected traffic loads on the links that share the same channel and that interfere with each other should not exceed the channel's raw capacity.

At first sight, channel assignment seems to be a straightforward problem of graph coloring [28]. However, standard graph coloring cannot capture the above constraints and specifications of the problem. A node-multi-coloring formulation [13] fails to capture the third constraint where the communicating nodes need a common color. On the other hand, an edge-coloring formulation fails to capture the second constraint where no more than the number of radios per node colors can be incident to a node. Although constrained edge-coloring might be able to roughly model the remaining constraints, it cannot satisfy the fourth constraint of limited channel capacity.

Moreover, a key problem in the design of channel assignment for multi-radio WMNs is the *channel dependency* among the logical links that share a common channel. Consider the WMN shown in Fig. 5.7 where six non-overlapping channels are available. Notice that links (***a,e***), (***e,d***), (***d,i***) and (***i,h***) all share channel 3 and therefore, if any of the nodes ***a***, ***e***, ***d***, ***i***, or ***h*** decides to reassign the channel on these virtual links, then the rest of the links have to change their assignment which then produces a *ripple effect*. This channel dependency among the nodes makes it difficult to predict the effect of node revisits or re-assignment.

Finally, a channel assignment algorithm should take into consideration the amount of traffic load on the virtual links. It may be assumed that each virtual link in the network has the same traffic load. However this does not hold true in most cases as some links generally carry more traffic than others [28] (for example, links associated with the gateway node). Generally speaking more bandwidth should be given to nodes that support higher traffic. In other words, channels assigned to these links should be shared among fewer nodes. Such traffic-aware channel assignment strategy would distribute the radio resources so as to match the distribution of traffic load in the mesh backbone.

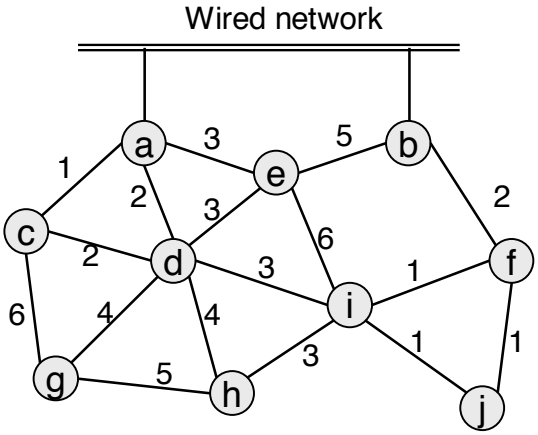



**Fig. 5.7.** Channel dependency.

Because channel assignment depends on the expected load on each virtual link, which in turn depends on routing, a circular dependency there exists between channel assignment and routing [28]. Routing depends on the capacity of virtual links, which is determined by channel assignment. This is because the capacity of a virtual link depends on the number of other links that are within its interference range and that are using the same channel. Similarly, channel assignment depends on the expected load of the virtual links, which is affected by routing. There are two different strategies to deal with this circularity between routing and channel assignment, as depicted in Fig. 5.8.

Given a set of node pairs and the expected traffic load between each node pair, according to the first strategy shown in Fig. 5.8(a), the routing algorithm devises the initial routes for the node pairs. Given these initial routes for the node pairs and hence the traffic load on each virtual link, the channel assignment algorithm assigns a channel to each radio taking into account the link traffic load. This assignment of channels is finally fed back to the routing algorithm. The second strategy, shown in Fig. 5.8(b), is different from the first in the sense that the routing algorithm assumes some initial assignment of channels to the radios. Based on this, the link capacities are estimated and passed to the routing algorithm, which in turn passes the link load needed for channel assignment. Obviously, both strategies may end up with inaccurate link capacities/link loads fed to the routing algorithm/channel assignment, which may require iterations between routing and channel assignment as in [28].

Some examples of the methods used for the estimation of link load and link capacity are presented in the next subsections.

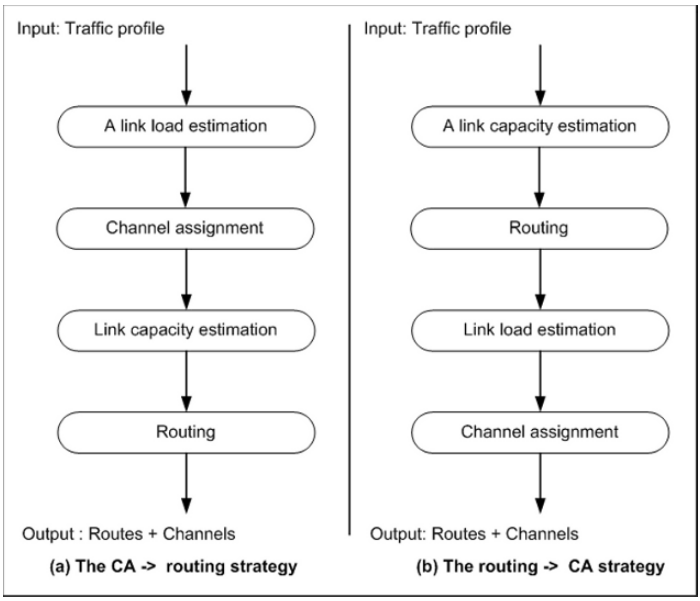



**Fig. 5.8.** Strategies for load aware channel assignment.

### 5.4.1 Link Load Estimation

There are several methods for deriving a rough estimate of the expected link traffic load. These methods depend on the routing strategy used (e.g. load balanced routing, multi-path routing, shortest path routing).

One approach is based on the concept of load criticality [8]. This method assumes perfect load balancing across all acceptable paths between each communicating pair of nodes. Let *P(s,d)* denote the number of acceptable paths (or virtual connections) between a pair of nodes (***s,d***), and let $P_l(s,d)$ be the number of acceptable paths between (***s,d***) that pass through a link ***l***. And finally, let *B(s,d)* be the estimated load between (***s,d***). Then the expected traffic load ($\Phi_l$) on link ***l*** is calculated as:

$$\Phi_l = \sum_{s,d} \frac{P_l(s,d)}{P(s,d)} \times B(s,d). \tag{5.1}$$

This equation implies that the initial expected traffic on a link is the sum of the loads from all acceptable paths, across all possible node pairs, which pass through the link. Because of the assumption of uniform multi-path routing, the load that an acceptable path between a pair of nodes is expected to carry is equal to the expected load of the pair of nodes divided by the total number of acceptable paths between them. Let us consider the same logical topology as shown in Fig. 5.7 and let us assume that we have the following three flows

Since we have three different sources and destinations, $\Phi_l$ will be equal to:

**Table 5.1.** Traffic profile with 3 flows.

| Source (s) | Destination (d) | B(s,d) (Mbps) |
|---|---|---|
| a | g | 0.9 |
| i | a | 1.2 |
| b | j | 0.5 |

$$\frac{P_l(a,g)}{P(a,g)} \times B(a,g) + \frac{P_l(i,a)}{P(i,a)} \times B(i,a) + \frac{P_l(b,j)}{P(b,j)} \times B(b,j). \qquad (5.2)$$

Furthermore, for each flow, let us assume the following are all the possible paths from source to destination. Consequently, we can also calculate *P(s,d)* for each flow.

**Table 5.2.** Possible flows between communicating nodes.

| (s,d) | (a,g) | (i,a) | (b,j) |
|---|---|---|---|
| Possible paths | a-c-g | i-e-a | b-f-j |
| | a-c-d-g | i-e-d-a | b-f-i-j |
| | a-d-g | i-d-a | b-e-i-j |
| | a-d-c-g | i-d-c-a | b-e-i-f-j |
| | a-d-h-g | i-d-e-a | b-e-d-i-j |
| | a-d-i-h-g | i-d-g-c-a | |
| | a-e-d-g | i-h-d-a | |
| | a-e-i-h-g | i-h-g-c-a | |
| P(s,d) | 8 | 8 | 5 |

From the above information, we can now calculate how many paths pass a specific link in the network topology. These values and the corresponding link traffic load ($\Phi_l$) calculated using equation (5.6) are shown in the following table

Based on these calculations, we can estimate the load between each neighboring node. The meaning of $\Phi_l$, which we have calculated throughout this example, is the link expected traffic load, i.e. the amount of traffic expected to be carried over a specific link. This representation of traffic between neighboring nodes is also referred to as the *traffic matrix*. The traffic matrix is an important estimate that enables a traffic aware channel assignment to be achieved.

### 5.4.2 Link Capacity Estimation

The link capacity, or the portion of channel bandwidth available to a virtual link, is determined by the number of all virtual links in its interference range that are also assigned to the same channel. Obviously, the exact short-term instantaneous bandwidth available to each link is dynamic and continuously changing depending on such complex system dynamics as physical obstacles, distance, capture effect,

**Table 5.3.** Possible flows between communicating nodes.

| Link ID l | $P_l(a,g)$ | $P_l(i,a)$ | $P_l(b,j)$ | $\Phi_l(Mbps)$ |
|---|---|---|---|---|
| 1 a-c | 2 | 3 | 0 | 0.675 |
| 2 c-g | 2 | 2 | 0 | 0.525 |
| 3 c-d | 2 | 1 | 0 | 0.375 |
| 4 d-g | 2 | 1 | 0 | 0.375 |
| 5 a-d | 4 | 3 | 0 | 0.9 |
| 6 g-h | 0 | 1 | 0 | 0.15 |
| 7 d-h | 1 | 1 | 0 | 0.2625 |
| 8 a-e | 2 | 2 | 0 | 0.525 |
| 9 d-e | 1 | 2 | 1 | 0.5125 |
| 10 d-i | 1 | 3 | 1 | 0.6625 |
| 11 h-i | 2 | 2 | 0 | 0.525 |
| 12 e-i | 1 | 2 | 2 | 0.6125 |
| 13 b-e | 0 | 0 | 3 | 0.3 |
| 14 b-f | 0 | 0 | 2 | 0.2 |
| 15 f-i | 0 | 0 | 2 | 0.2 |
| 16 i-j | 0 | 0 | 2 | 0.2 |
| 17 f-j | 0 | 0 | 2 | 0.2 |

coherence period, and stray radio frequency (RF) interferences [28]. The goal here is to derive an approximation of the long-term bandwidth share available to a virtual link. One approximation of a virtual link ***i***'s capacity $bw_i$ can be obtained using the following equation:

$$bw_i = \frac{\Phi_i}{\sum_{j \in Intf(i)} \Phi_j} \times C \tag{5.3}$$

where $\Phi_i$ is the expected load on link ***i***, *Intf(i)* is the set of all virtual links in the interference zone of link ***i***, and ***C*** is the sustained radio channel capacity. The rationale behind this formula is that when a channel is not overloaded, the channel share available to a virtual link is proportional to its expected load. The higher the expected load on a link, the more channel share it should get. The accuracy of this formula decreases as $\sum_{j \in Intf(i)} \Phi_j$ approaches ***C***.

To summarize, the inputs to a channel assignment algorithm are: (1) the connectivity graph, (2) the number of non-overlapping channels, (3) the number of radios available on each mesh router, and (4) an estimated traffic load for each communicating pair of nodes. The output is the channel bound to each radio in the multi-radio WMN.

In the next section, we will present various channel assignment schemes proposed in the literature.

## 5.5 Taxonomy of Channel Assignment Schemes for WMNs

As has been already mentioned, Channel Assignment (CA) in a multi-radio WMN environment consists of assigning channels to the radios in order to achieve efficient channel utilization (i.e. minimize interference) and, at the same time, to guarantee an adequate level of connectivity. The problem of optimally assigning channels in an arbitrary mesh topology has been proven to be NP-hard based on its mapping to a *graph-coloring* problem [28]. Therefore, channel assignment schemes predominantly employ heuristic techniques to assign channels to radios belonging to WMN nodes. In this section, we present a taxonomical classification of various CA schemes for mesh networks. Fig. 5.9 presents the taxonomy on which the rest of the section is based. Specifically, the proposed CA schemes can be partitioned into three main categories - *fixed, dynamic* and *hybrid* - depending on how frequently the CA scheme is modified. In a fixed scheme the CA is almost constant, while in a dynamic scheme it is continuously updated to improve performance. A hybrid scheme applies a fixed scheme for some radios and a dynamic one for others. We will now analyze these three categories and give examples of CA schemes from each category.

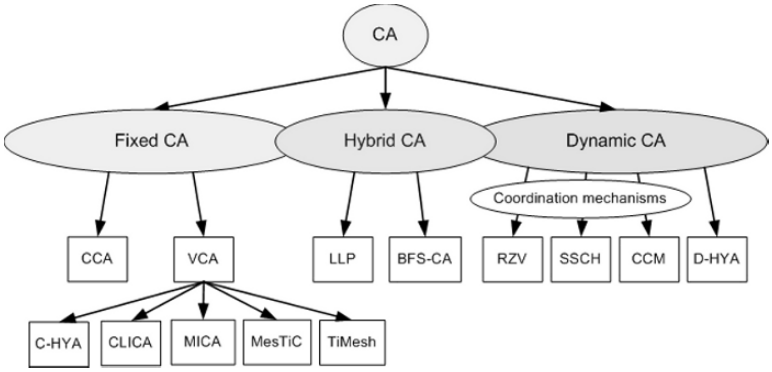



**Fig. 5.9.** Taxonomy of channel assignment schemes in wireless mesh networks.

### 5.5.1 Fixed Channel Assignment Schemes

Fixed assignment schemes assign channels to radios either permanently, or for time intervals that are long with respect to the radio switching time. Such schemes can be further subdivided into *common channel assignment* and *varying channel assignment*.

#### Common Channel Assignment (CCA)

This is the simplest scheme. In CCA [6], the radios of each node are all assigned the same set of channels. For example, if each node has two radios, then the same

two channels are used at every node as shown in Fig. 5.10. The main benefit is that the connectivity of the network is the same as that of a single channel approach, while the use of multiple channels increases network throughput. However, the gain may be limited in scenarios where the number of non-overlapping channels is much greater than the number of radios available in each node. Thus, although this scheme presents a simple CA strategy, it does not take into account all the various factors affecting the performance of a channel assignment in a WMN, thus producing an inefficient utilization of the network resources (i.e. interference).

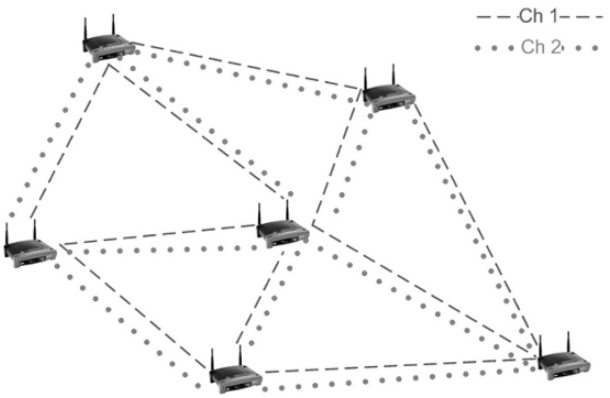



**Fig. 5.10.** An example of common channel assignment.

**Varying Channel Assignment (VCA)**

In the VCA scheme, radios of different nodes may be assigned different sets of channels [21, 28]. However, the assignment of channels may lead to network partitions and/or topology changes, which may increase the length of routes between mesh nodes. Therefore, in this scheme, channel assignment needs to be carried out carefully. Below we discuss the VCA approach in more details by presenting five algorithms that belong to this sub-category.

*Centralized Channel Assignment (C-HYA)*

Based on Hyacinth, a multi-channel wireless mesh network architecture, a centralized channel assignment algorithm for WMNs was proposed in [28], where traffic is mainly directed toward gateway nodes, i.e. the traffic is directed to/from the Internet. Assuming that the offered traffic load is known, this algorithm assigns channels thus ensuring network connectivity and satisfying the bandwidth limitations of each link. It first estimates the total expected load on each virtual link by summing the load due to each offered traffic flow. Then, the channel assignment algorithm visits each

virtual link in decreasing order of expected traffic load and greedily assigns it a channel. The algorithm starts with an initial estimation of the expected traffic load and iterates over channel assignment and routing until the bandwidth allocated to each virtual link matches its expected load. Although this scheme presents a method for channel allocation that incorporates connectivity and traffic patterns, the assignment of channels on links may cause a *ripple effect* (see Section 5.4) whereby already assigned links have to be revisited, thus increasing the time complexity of the scheme. An example of node revisiting is illustrated in Fig. 5.11. In this case, node ***a*** is assigned channels 1 and 4, and node ***b*** channels 3 and 8. Because ***a*** and ***b*** have no common channel, a channel re-assignment is required. Specifically, link (***a,b***) needs to be assigned one of the channels from [1, 3, 4, 8]. Based on the channel expected loads, link (***a,b***) is assigned channel 1, and channel 8 assigned already to link (***b,d***) is changed to channel 1.

The results in [28] showed that by deploying only two radios per node, it is possible to achieve a factor of up to 8 in the improvement of the overall network goodput when compared to the case of a single-radio inherently limited to a single channel.

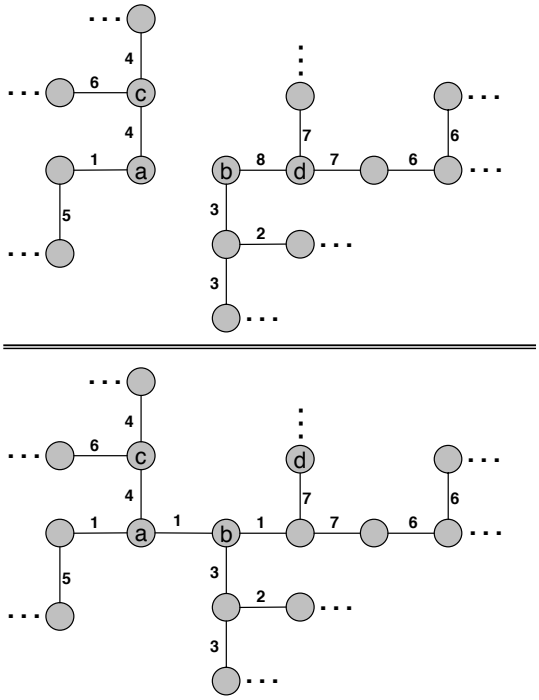



**Fig. 5.11.** An example of channel revisit in C-HYA.

*A Topology Control Approach (CLICA)*

A polynomial time greedy heuristic, called *Connected Low Interference Channel Assignment (CLICA)*, was presented in [21] to enable an efficient and flexible topology formation, ease of coordination, and to exploit the static nature of mesh routers to update the channel assignment on large timescales.

CLICA is a traffic independent channel assignment scheme which computes the priority for each mesh node and assigns channels based on the connectivity graph and on the conflict graph. However, the algorithm can override the priority of a node to account for the lack of flexibility in terms of channel assignment and to ensure network connectivity. Although this scheme avoids link revisits, it does not incorporate the role of traffic patterns (an example of traffic pattern is shown in Table 5.1) in channel assignment for WMNs.

To understand the functioning of the CLICA algorithm, let us consider the example in Fig. 5.12. Suppose nodes ***a*** and ***d*** have two radios and the initial order of priorities is ***a***, ***d***, ***c*** and ***b***. CLICA starts at ***a*** to color its incident links; it starts by coloring link (***a,b***) with channel *C1*. As a result, ***b*** loses further flexibility in choosing channels for its other incident links. So, CLICA bumps ***b***'s priority to the highest. Moreover, it recursively starts assigning channels at ***b***, which results in node ***b*** reusing channel *C1* for link (***b,c***). The same procedure as above (i.e., priority increase followed by recursive color reuse) is repeated at node ***c*** thus forcing link (***c,d***) to use *C1*. Now, because ***d*** has two radios and only one of them is already assigned, the algorithm assigns link (***a,d***) with *C2* by using the additional radios.

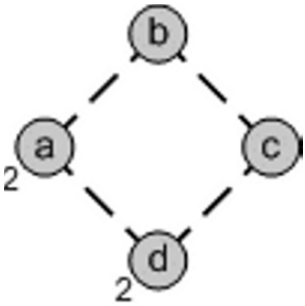



**Fig. 5.12.** Connectivity graph.

Note that, CLICA is naturally recursive and follows a chain of the least flexible nodes to maintain network connectivity. Also note that it is a one-pass algorithm in the sense that once coloring decisions have been made, they are not reversed later in the algorithm execution. Simulation results demonstrate the effectiveness of CLICA in reducing interference, which represents the objective function for the CA optimization problem.

*Minimum-Interference Channel Assignment (MICA)*

In [35] the authors extended [21] and developed two new algorithms. The first is based on a popular heuristic search technique called Tabu search [9], which was originally designed for graph coloring problems. The second is a greedy heuristic inspired by the greedy approximation algorithm for Max K-cut [7] problem in graphs. The Tabu-search based method starts with a random assignment. A neighborhood search is then run for a better solution by flipping the assignment of some nodes. At the same time, the method remembers the best solution seen so far and stops when the maximum number of iterations allowed is reached without a better solution being found (an example of an output of the first phase is shown in Fig. 13(a)). This solution is the best without taking into account the interface constraint, i.e. the total number of available channels at any network node is less than or equal to the number of radios on that node. Therefore, the last step in the algorithm is to start from the node with the maximum violations of the interface constraint, and combine any assignments of radios that share the same channel and share an edge between them in such a way as to minimize the increase in conflicts.

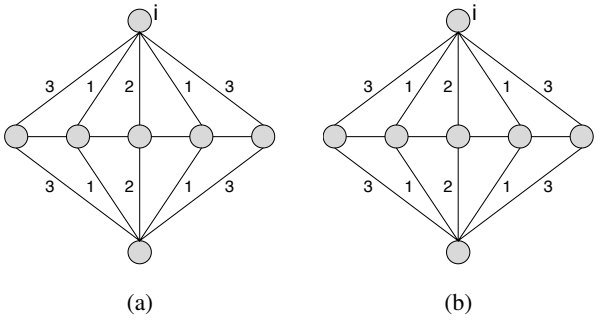



**Fig. 5.13.** Merge operation of second phase: (a) output before the second phase and (b) output after the second phase.

In Fig. 5.13(a) ***i*** stands for the node picked for the merge operation. The number of colors incident on ***i*** is reduced by picking two colors *C3* and *C2* that are incident on ***i***, and changing the color of all *C3*-colored links to *C2*. In order to ensure that this change does not create interface constraint violations on other nodes, the change will iteratively propagate to all *C3*-colored links that are connected to the links whose color has been changed from *C1* to *C2* (two links are said to be connected if they are incident on a common node). Essentially, the above propagation of color change ensures that for any node ***j***, either all or none of the *C3*-colored links incident on ***j*** are changed to color *C2*. The result of the merge operation after the second phase is shown in Fig. 5.13(b).

On the other hand, the second greedy heuristic developed in [35], based on Max K-cut, takes care of the interface constraint at each iteration. The Max K-cut problem

consists of how to partition the vertex set of a graph into k sets so as to maximize the number of edges crossing between partitions. Using linear programming and semi-definite programming formulations of this optimization problem, tight lower bounds on the optimal network interference was obtained.

*Traffic and Interference Aware Channel Assignment Scheme (MesTiC)*

MesTiC [31] stands for ***Mes***h based ***T***raffic and ***i***nterference aware ***C***hannel assignment. It is a fixed, rank-based, polynomial time greedy algorithm for centralized CA, which visits nodes once in the decreasing order of their rank. The rank of each node ***R*** is computed on the basis of its link traffic characteristics, topological properties and number of radios on a node according to the following ratio:

$$\boldsymbol{R}\ (node) = \frac{Aggregate\ traffic\ (node)}{min\ hops\ from\ gateway\ (node) * number\ of\ radios\ (node).} \tag{5.4}$$

Clearly, the aggregate traffic flowing through a mesh node has an impact on the channel assignment strategy. The rationale is that if a node relays more traffic, assigning it a channel of least interference will increase the network throughput. Thus, aggregate traffic in the numerator in equation 5.2 increases the rank of a node with its traffic. In addition, due to the hierarchical nature of a mesh topology, the nodes nearest to the gateway should have a higher preference (rank) in channel assignment, as they are more likely to carry more traffic. At the same time, the number of radios on a node gives flexibility in channel assignments and should inversely affect its priority (i.e. the lower the number of radios, the higher the priority in channel assignment).

MesTiC ensures the topological connectivity by using a common default channel deployed on a separate radio on each node, which can also be used for network management purposes. Fixed schemes alleviate the need for channel switching, especially when switching delays are large, as is the case with the current 802.11 hardware. In addition, MesTiC is rank-based, which gives the nodes that are expected to carry heavy loads more flexibility in assigning channels. Finally, the use of a common default channel prevents flow disruption, as discussed in [26].

The MesTiC algorithm traverses the mesh network nodes in descending order of their rank assigning channels to the radios. For further details on MesTiC see [31] and [32].

Let us illustrate the working principle of MesTiC by considering the simple example in Fig. 5.14(a) where the input connectivity graph and estimated link traffic (i.e. the estimated traffic between a node and its neighbors) are shown. The network is configured with three channels and two radios per node. Assuming that node ***b*** is the gateway node, the rank of the remaining nodes, in decreasing order, is ***d***, ***a***, ***c***. The algorithm starts by visiting node ***b*** first, assigning channel *C1* to the link between (***b,a***) (which carries the highest traffic of 120), and then moves on to assign channel *C2* to the link (***b,d***). Now, when assigning a channel to link (***b,***c), it has to choose between *C1* and *C2*. However, as *C1* carries more traffic than *C2*, it assigns *C2* to link (***b,c***). Likewise, at node ***d***, it assigns a previously unassigned channel *C3* to the link

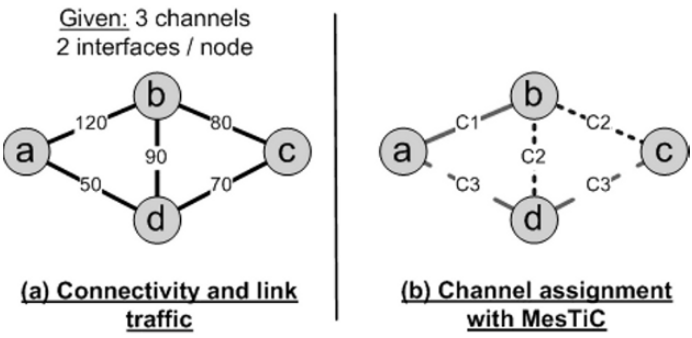



**Fig. 5.14.** An example illustrating how MesTiC works.

(***d,c***) and, as *C3* carries less traffic than *C2* (90 + 80 =170) or *C1* (120), it assigns *C3* to the link (***d,a***). The algorithm proceeds until all links and radios have been assigned channels, as shown in Fig. 5.14(b). Simulation results show that MesTiC performs better than other CA algorithms for several topologies and traffic profiles.

*Topology Design and Channel Assignment (TiMesh)*

In [25], the authors presented a decentralized channel assignment strategy that considers topology control and channel allocation as two separate but related problems. The former takes care of channel dependency (see Section 5.4) and the latter deals interference. The logical topology formation and radio assignment are formulated as a joint optimization problem based on a Multi-channel WMN (MC-WMN) architecture called TiMesh. The model of the proposed solution takes into account: the number of radios on each mesh router, the channel dependency among the nodes that share a common channel, the degree of a node, and the expected traffic load between the various source and destination nodes. The goals are: (1) to guarantee network connectivity, by supporting both internal traffic (among the wireless routers) and external traffic (to the internet); (2) to prevent ripple effects among the logical links sharing the same channel.

The MC-WMN is modeled by a physical topology graph G(N,E). Where N is the set of mesh routers (each equipped with ***I*** radios) and ***E*** is the set of links between the mesh routers.

The first constraint to the problem is that logical links are assumed to be bidirectional. The second constraint is channel dependency. To restrict this dependency an upper bound on the number of additional logical links that may share a radio with a particular link is set. The larger this value is, the smaller the proportion of time that each logical link can access the shared radio. The third constraint is the ripple effect. The approach is to assign an exclusive radio to one end of each logical link. This means that if node ***x*** is responsible for the channel allocation on logical link (***x,y***), then the radio that is assigned by node ***y*** to attach to link (***x,y***) should not be used by any other logical link. For capacity planning, a statistical model of the network traffic

is used and flow conservation is applied at each node. This guarantees that there is at least one path available between each source and destination pair (***s,d***). Thus, the obtained topology is always connected.

The fourth constraint is the hop count, which states that for each source and destination pair (***s,d***), there exists at least one path where the hop count is less than or equal to the shortest path + a tunable parameter $\Gamma$ (a positive integer).

It is assumed that a power control algorithm maintains a constant data rate in the presence of fading and other channel imperfections. This implies that there is a fixed nominal capacity associated with the logical links. However, the actual capacity depends on the number of additional logical links that are sharing the same channel. The utilization of the logical link is then defined as the total traffic load between source and destination (which is assumed to be known) divided by the effective link capacity.

The objective function for the optimization problem is to minimize the maximum utilization across all the links given the constraints defined earlier. For TiMesh, a fast greedy algorithm [22] was used to provide the solutions for the logical topology design and radio assignment problems. Moreover, the solution also determines which end node on each logical link is responsible for channel allocation.

### 5.5.2 Dynamic Channel Assignment Schemes

As in the fixed CA, dynamic CA strategies allow any radio to be assigned any channel but in the latter CA radios can frequently switch from one channel to another. Therefore, when nodes need to communicate with each other, in a dynamic CA, a coordination mechanism has to ensure that they are on a common channel. For example, the coordination mechanism may require all nodes to visit a predetermined "rendezvous" channel [33] periodically to negotiate channels for the next phase of transmissions as shown in Fig. 5.15.

Another mechanism, called *Slotted Seeded Channel Hopping (SSCH)*, consists in using pseudo-random sequences [2] where each node should switch channels synchronously in a pseudo-random sequence so that all neighbors meet periodically in the same channel. In this approach the interfaces must be capable of fast synchronous channel switching. Specifically, time is divided into slots and the channels are switched at the beginning of each slot according to:

$$New\ Channel\ =\ (Old\ Channel\ +\ Seed)\ mod\ (Number\ of\ Channels). \tag{5.5}$$

An example of the SSCH mechanism is illustrated in Fig. 5.16.

Another approach to dynamic channel assignment is the control channel approach, shown in Fig. 5.17, where one radio is assigned to a common channel for control purposes, and the rest of the radios are switched between the remaining channels and used for data exchange [36]. The benefit of dynamic assignment is the ability to switch a radio to any channel, thereby offering the potential of using many channels with only a few radios. The key challenge with the dynamic switching approach

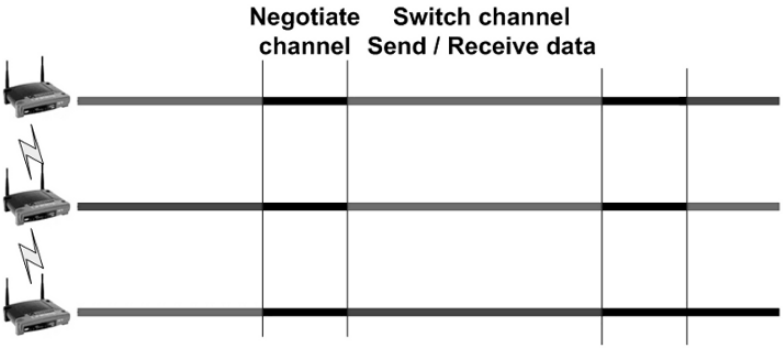



**Fig. 5.15.** An example of the synchronization "rendezvous" mechanism.

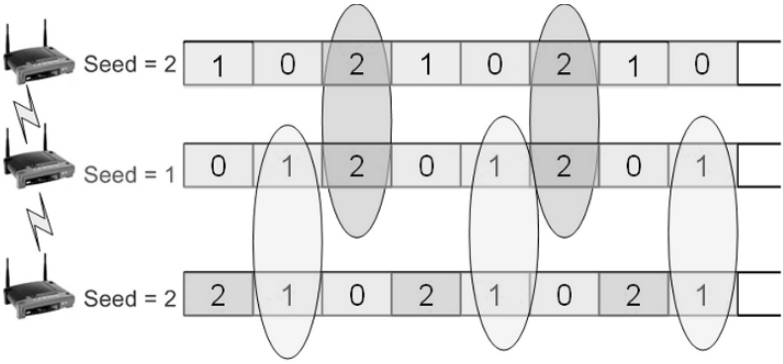



**Fig. 5.16.** Example of SSCH: Slotted Seeded Channel Hopping.

is how to coordinate the decisions in terms of when to switch radios, as well as what channel to switch the radios to.

### A Distributed Channel Assignment Scheme (D-HYA)

A set of dynamic and distributed channel assignment algorithms was proposed in [29, 30]. These algorithms can react to traffic load changes in order to improve the aggregate throughput and achieve load balancing. Based on the *Hyacinth* architecture, the algorithm (described in [29] as well as in [30] with a slight change) builds on a spanning tree network topology, similar in construction to that of IEEE 802.1D. The scheme works in such a way that each gateway node is the root of a spanning tree, and every mesh node belongs to one of these trees. The channel assignment problem consists of the following two steps.

(a) *Neighbor-to-interface binding* (i.e. the node selects the radio to communicate with every neighbor), where dependency among the nodes is eliminated in order to prevent *ripple effects* in the network [28]. This is achieved by imposing a restriction

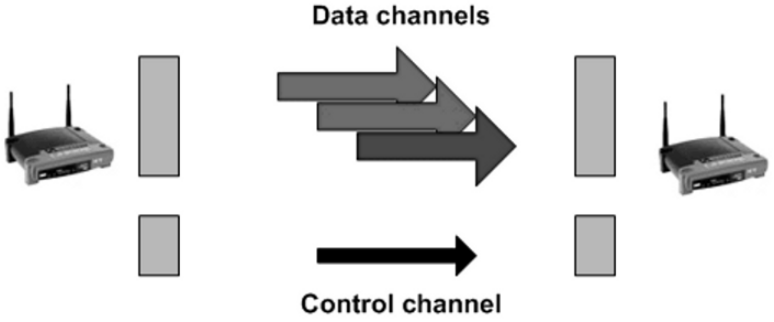



**Fig. 5.17.** An example of the control channel mechanism.

that the set of radios that a node uses to communicate with its parent node, termed *UP-NICs*, is disjoint from the set of radios the node uses to communicate with its children nodes, called *DOWN-NICs*.

(b) *Interface-to-channel binding* (i.e. the node selects the channel to assign to every radio), where the goal is to balance the load among the nodes and relieve interference. The channel assignment of a WMN node's *UP-NICs* is the responsibility of its parent. To assign channels to *DOWN-NICs*, a WMN node needs to estimate the usage status of all the channels within its interference neighborhood. Each node therefore periodically exchanges its individual channel usage information as a *CHNL USAGE* packet with all its neighbors. Based on the per-channel total load information, a WMN node determines a set of channels that are least used in its vicinity. As nodes higher up in the spanning trees need more relay bandwidth, they are given a higher priority in channel assignment. More specifically, the priority of a WMN node is equal to its hop distance from the gateway.

When a WMN node performs channel assignment, it restricts its search to the channels that are not used by any of its interfering neighbors with a higher priority. The outcome of this priority mechanism is a fat-tree architecture where links higher up in the tree are given higher bandwidth. Because traffic patterns and thus channel loads can evolve over time, radio-to-channel mapping is adjusted periodically, every Tc time units. Within a channel load-balancing phase, a WMN node evaluates its current channel assignment based on the channel usage information it receives from neighboring nodes. As soon as the node finds a relatively less loaded channel after accounting for priority and its own usage of the current channel, it moves one of its *DOWN-NICs* operating on a heavily-loaded channel to use the less-loaded channel. It also sends a *CHNL CHANGE* message with the new channel information to the affected child nodes, which modify the channels of their *UP-NICs* accordingly.

To summarize, in D-HYA channels are dynamically assigned to the radios based on their traffic load. However, the tree-topology constraint of the scheme poses a potential hindrance in leveraging multi-path routing in mesh networks.

### 5.5.3 Hybrid Channel Assignment Schemes

Hybrid channel assignment strategies combine both static and dynamic assignment properties by applying a fixed assignment for some radios and a dynamic assignment for other radios (see for example [17, 18, 26]). Hybrid strategies can be further classified based on whether the fixed radios use a common channel [26] or a varying channel [17, 18] approach. The fixed radios can be assigned a dedicated control channel [29] or a data and control channel [26], on the other hand the other radios can be switched dynamically among channels. Hybrid assignment strategies are attractive because, as with fixed assignment, they allow for simple coordination algorithms, while still retaining the flexibility of dynamic channel assignment. In the next two sub-sections, we will describe two hybrid CA schemes.

#### Link Layer Protocols for Radio Assignment (LLP)

In [17, 18], an innovative link layer radio assignment algorithm was proposed that categorizes available radios into fixed (F) and switchable (S) radios. Fixed radios are assigned, for long time intervals, specific fixed channels, which can be different for different nodes. On the other hand, switchable radios can be switched over short time scales among the non-fixed channels based on the amount of data traffic. By distributing fixed radios of different nodes on different channels, all channels can be used, while the switchable radio can be used to maintain connectivity. Fig. 5.18 illustrates how the protocol works where node *A*, *B* and *C*'s fixed radios are assigned channels 3, 2 and 1 respectively. Now assume node *B* wishes to exchange data with nodes *A* and *C*. When *B* has to send a packet to *A*, *B* switches its switchable radio to channel 3 and transmits the packet. Since *A* is always listening to channel 3 with its fixed radio, *A* can receive the transmission of *B*. Now if *A* has to send a packet back to *B*, *A* switches its switchable radio to channel 2 and transmits the packet. Since *B* is listening to channel 2 with its fixed radio, the packet from *A* can be received. Similarly, if *B* has to subsequently send a packet to *C*, it switches to channel 1 and sends the packet. Note that *A* and *C* can at any time send a packet to *A* on channel 2. Thus, there is no need to coordinate when to schedule transmissions among *A*, *B*, and *C*.

Two coordination protocols were proposed in [17] to decide which channels should be assigned to the fixed radio, and to manage communication between the nodes. The first is the use of a well-known function that generates a hash based on the node identifier to select which channel to assign to the fixed radio. Neighbors of this node can use the same function to compute which channel to use to communicate with this node. The second strategy is the explicit exchange of *Hello* packets that contain information on the fixed channel used by a node. Based on the received *Hello* packets, nodes may (with some probability, to avoid oscillations) choose to set their fixed channel to an unused or a lightly loaded channel.

In [18], the authors proposed a hybrid CA scheme based on the second coordination protocol, which works as follows. Periodically, each node broadcasts a *Hello* packet on every channel. The *Hello* packet contains the fixed channel being used by

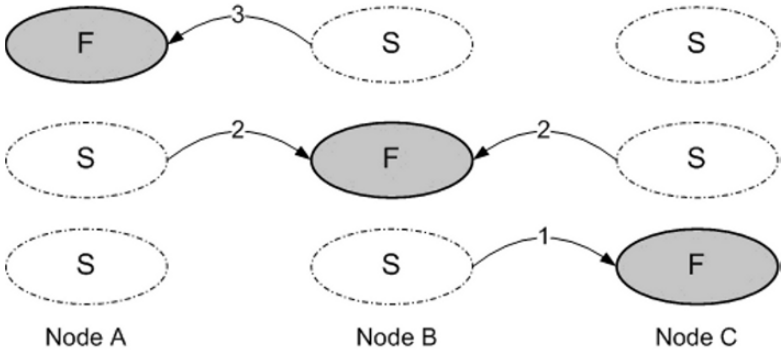



**Fig. 5.18.** Hybrid protocol operation.

the node, and its current *NeighborTable*. When a node receives a *Hello* packet from a neighbor, it updates its *NeighborTable* with the fixed channel of that neighbor. The *ChannelUsageList* is updated using the *NeighborTable* of its neighbor. Updating *ChannelUsageList* with each neighbor's *NeighborTable* ensures that the *ChannelUsageList* will contain two-hop channel usage information. An entry that has not been updated for a specified maximum lifetime is removed. This ensures that stale entries of nodes that have moved away are removed from the *NeighborTable* and *ChannelUsageList*.

The main benefit of this hybrid protocol is that it is fairly insensitive to radio switching delay, however, the assignment of fixed channels has to be carefully balanced in order to achieve a good performance.

### Interference-Aware Channel Assignment (BFS-CA)

The channel assignment problem in WMNs in the presence of interference from colocated wireless networks was addressed in [26]. The authors proposed a dynamic, centralized, interference-aware algorithm aimed at improving the capacity of the WMN backbone and at minimizing interference. This algorithm is based on an extension to the conflict graph concept called the multi-radio conflict graph (MCG) (Section 5.3.3) where the vertices in the MCG represent edges between radios instead of edges between mesh routers. To compensate for the drawbacks of a dynamic network topology, the proposed solution assigns one radio on each node to operate on a default common channel throughout the network. This strategy ensures a common network connectivity graph, provides alternate fallback routes and avoids flow disruption by traffic redirection over a default channel. This scheme computes interference and bandwidth estimates based on the number of *interfering radios*, where an interfering radio is a simultaneously operating radio that is visible to a mesh router but is external to its network. Moreover, a measurement of only the number of interfering radios is not considered sufficient because it does not indicate the amount of traffic generated by the interfering radios. For instance, two channels could have the

same number of interfering radios but one channel may be more heavily used by the interfering radios compared to the other.

Therefore, each mesh router also estimates the bandwidth used by the interfering radios. Each mesh router then derives two separate channel rankings. The first ranking depends on the increasing number of interfering radios. The second depends on the increasing channel utilization. The mesh router then merges the two rankings by taking the average of the individual ranks. The resulting ranking is used by the CA scheme. This scheme, called the *Breadth First Search Channel Assignment (BFS-CA)* algorithm, uses a breadth first search to assign channels to the radios. The search begins with links emanating from the gateway node; while links fanning outwards towards the edge of the network are given lower priority.

The default channel is chosen so that its use in the mesh network minimizes interference between the mesh network and collocated wireless networks. This is achieved by computing the rank $R_c$ of a channel as follows:

$$R_c = \frac{\sum_{i=1}^{n} Rank_c^i}{n} \tag{5.6}$$

where $\boldsymbol{n}$ is the number of routers in the mesh and $Rank_c^i$ is the rank of channel $\boldsymbol{c}$ at router $\boldsymbol{i}$. The default channel is then chosen as the channel with the least $R_c$ value.

The assignment of non-default channels, on the other hand, is based on information in the MCG where it is associated with every vertex its corresponding link delay value computed based on the Expected Transmission Time or ETT [6]. The CA scheme also associates with each vertex a channel ranking derived by taking the average of the individual channel rankings of the two radios that make up the vertex. The average is important because the assignment of a channel to a vertex in the MCG should take into account the preferences of both end-point radios that make up the vertex. Once the channel assignments have been decided, the mesh routers are notified to re-assign their radios to the chosen channels as described in detail in [26]. To adapt to the changing interference characteristics, the CA periodically re-assigns channels. The periodicity depends ultimately on how frequently interference levels in the mesh network are expected to change.

## 5.6 Comparisons of CA Schemes

The most important features of the existing CA algorithms for WMNs are summarized in Fig 5.19 The key issues are: connectivity, topology control, interference minimization and traffic pattern. C-HYA is a traffic-aware CA scheme. While its distributed version, D-HYA, alleviates the effect of link revisits, stringent restrictions were imposed on the topology of the mesh network, thereby failing to leverage the advantages of multi-path routing in a mesh scenario. MesTiC is a fixed, centralized scheme that in the same way as C-HYA and D-HYA take traffic load information into account, without, at the same time imposing any strong constraints on the topology. Moreover, it is a greedy algorithm, which does not suffer from ripple effects and ensures connectivity via a default radio. Although the goal of LLP and CLICA

was to minimize interference, the effect of traffic patterns on interference and thus on the CA scheme, was not taken into account. The effect of traffic in BFS-CA was considered, but only for traffic emanating from external wireless networks. From another perspective, some algorithms, such as CLICA, MICA and TiMesh considered topology control, which incurs overheads in the channel assignment algorithm but alleviates the need for an additional radio tuned to a common channel. On the other hand, others (e.g. BFS-CA, MesTiC) assume default connectivity by using a separate common channel on a separate radio.

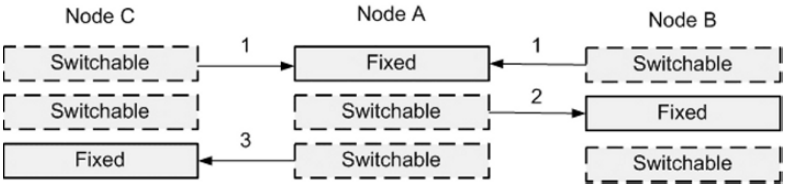



**Fig. 5.19.** Comparative study of the salient features of channel assignment schemes.

## Conclusion

In this chapter we have identified the key challenges associated with assigning channels to radio interfaces in a multi-radio wireless mesh network. After presenting the channel assignment problem and its major constraints, we have provided a taxonomy of existing channel assignment schemes and summarized this survey with a comparison of the different schemes. One of the important challenges still to be solved is the question of how many interfaces to have on each mesh router. In other words, given the physical topology and the traffic profile of the network, how can we optimize the number of radios on the different nodes. This question adds another dimension to the channel assignment problem and still needs future investigation. Another important challenge arises when the nature of traffic is not uniform; for example, in a case when there is a mixture of broadcast, multicast and unicast traffic in the same network. This problem was discussed in [24] where the authors investigated extensively the channel assignment problem in the broadcast case. They discovered that for broadcast, a *common channel assignment* generally performs better than *variable channel assignment*. On the other end, CCA performs poorly for unicast flows and thus the challenge is to discover what channel assignment schemes can perform well for both.

## References

1. I. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," in *Computer Networks'05*, vol. 47, no. 47, pp. 445-487.

2. P. Bahl, R. Chandra, and J. Dunagan,"SSCH: Slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks," in *Proc. ACM MobiCom*, pp. 216-230, Feb. 2004.
3. R. Bruno, M. Conti, and M. Gregori, "Mesh networks: Commodity multi-hop ad hoc networks," *IEEE Communications Magazine*, pp. 123-131, 2005.
4. R. Choudhury, X. ChoudhuryYang, R. Ramanathan, and N. Vaidya, "Using directional antennas for medium access control in ad hoc networks," in *Proc. ACM MobiCom*, pp. 59-70, 2002.
5. J. Deng, B. Liang, and P. K. Varshney, "Tuning the carrier sensing range of IEEE 802.11 MAC," in *Proc. IEEE Global Telecommunications Conference'04*, vol. 5, pp. 2987-2991, 2004.
6. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Algoritmica, 18*, 1997.
7. A. Frieze and A. Jerrum, "Improved approximation algorithms for MAX k-CUT and MAX BISECTION," in *Proc. IEEE WoWMoM'06*, pp. 10, June 2006.
8. K. Gopalan, "Efficient network resource allocation with QoS guarantees," in *TR n133, ECSL, SUNY-SB*.
9. A. Hertz and D. de Werra, "Using Tabu search techniques for graph coloring," in *Computing, 39(4)*, 1987.
10. "IEEE 802.11a Standard," in *http://standards.ieee.org/getieee802/download/802.11a-1999.pdf, 1999.*
11. "IEEE 802.11b Standard," in *http://standards.ieee.org/getieee802/download/802.11b-1999.pdf, 1999.*
12. K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qui, "Impact of interference on multi-hop wireless network performance," in *Proc. ACM MobiCom'03*, pp. 66-80, 2003.
13. T. R. Jensen and B. Toft, *Graph Coloring Problems*, Wiley Interscience, New York, 1995.
14. I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Personal Communications*, pp. 10-31, 1996.
15. Y. B. Ko, V. Shankarkumar, and N. H. Vaidya, "Medium access control protocols using directional antennas in ad hoc networks," in *Proc. IEEE Annual Conference on Computer Communications(INFOCOM)*, pp. 13-21, 2000.
16. M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM MobiCom'05*, pp. 73-87, 2005.
17. P. Kyasanur and N. Vaidya, "Routing and interface assignment in multi-channel multi-interface wireless networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 2051- 2056, 2005.
18. P. Kyasanur and N. Vaidya, "Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks," *Mobile Computing and Communications Review*, vol. 10, no. 1, pp. 31-43, 2006.
19. S. Liese, D. Wu, and P. Mohapatra, "Experimental characterization of an 802.11b wireless mesh network," in *Proc. ACM IWCMC'06*, pp. 587-592, 2006.
20. A. Lozano, F. R. Farrokhi, and R. A. Valenzuela, "Lifting the limits on high-speed wireless data access using antenna arrays," *IEEE Communications Magazine*, pp. 156-162, 2001.
21. M. Marina and S. R. Das, "A topology control approach for utilizing multiple channels in multi-radio wireless mesh networks," in *Proc. IEEE Broadnets'05*, pp. 381-390, 2005.
22. S. Martello and P. Toth, "Heuristic algorithms for the multiple knapsack problem," *Journal of Computing*, vol. 27, pp. 93-112, 1981.

23. A. Nasipuri, S. Ye, and R. E. Hiromoto, "A MAC protocol for mobile ad hoc networks using directional antennas," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1214-1219, 2000.
24. J. Qadir, A. Misra, and C. T. Chou, "Minimum latency broadcasting in multi-radio multi-channel multi-rate wireless meshes," in *Proc. 3rd IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pp. 80-89, 2006.
25. A. H. M. Rad and V. W. S. Wong, "Logical topology design and interface assignment for multi-channel wireless mesh networks," in *Proc. IEEE Global Telecommunications Conference (Globecom)*, 2006.
26. K. Ramachandran, K. Almeroth, E. Belding-Royer, and M. Buddhikot,"Interference aware channel assignment in multi-radio wireless mesh networks," in *Proc. IEEE Infocom*, 2006.
27. R. Ramanathan, J. Redi, C. Santivanez, and S. Polit, "Ad hoc networking with directional antennas: A complete system solution," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 375-380, 2004.
28. A. Raniwala, K. Gopalan, and T. Chiueh, "Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks,"*ACM Mobile Computing and Communications Review (MC2R)*, 2004.
29. A. Raniwala and T. Chiueh, "Evaluation of a wireless enterprise backbone network architecture," in *Proc. 12th Hot-Interconnects*, pp. 98-104, 2004.
30. A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. IEEE Infocom*, pp. 2223-2234, 2007.
31. H. Skalli, S. Ghosh, S. K. Das, L. Lenzini, and M. Conti, "Channel assignment strategies for multi-radio wireless mesh networks: Issues and solutions," *IEEE Communications Magazine*, Special Issue on "Wireless Mesh Networks", Nov. 2007.
32. H. Skalli, S. K. Das, L. Lenzini, and M. Conti, "Traffic and interference aware channel assignment for multi-radio wireless mesh networks," Technical Report, IMT Lucca, 2006, *http://www.imtlucca.it/_documents/publications/publication54-9369_Technical_report_2006.pdf*.
33. J. So and N. Vaidya, "Multi-channel MAC for ad hoc networks: Handling multi-channel hidden terminals using a single transceiver," in *Proc. ACM Mobihoc*, pp. 222-233, 2004.
34. A. Spyropoulos and C. S. Raghavendra, "Energy efficient communications in ad hoc networks using directional antenna," in *Proc. IEEE Annual Conference on Computer Communications (INFOCOM)*, pp. 220-228, 2002.
35. A. Subramanian, H. Gupta, and S. R. Das, "Minimum-interference channel assignment in multi-radio wireless mesh networks," *Student Poster Session, 13th International Conference on Network Protocols (ICNP)*, 2005.
36. S. Wu, C. Lin, Y. Tseng, and J. Sheu, "A new multi-channel MAC protocol with on-demand channel assignment for multi-hop mobile ad hoc networks," submitted to the *International Symposium on Parallel Architectures, Algorithms, and Networks (ISPAN)*, 2000.

# 6

# Optimal Resource Allocation for Wireless Mesh Networks

Y. Xue[1], Y. Cui[1], and K. Nahrstedt[2]

[1] Vanderbilt University, USA
{yuan.xue, yi.cui}@vanderbilt.edu
[2] University of Illinois at Urbana-Champaign, USA
klara@cs.uiuc.edu

## 6.1 Introduction

Wireless networks enable ubiquitous information and computational resource access, and become a popular networking solution. Recently, wireless mesh networks (WMN) [1]- [7] have attracted increasing attention and deployment as a high-performance and low-cost solution to last-mile broadband Internet access.

In this chapter, we study the problem of *resource allocation* in wireless mesh networks. Our goal is to design effective resource allocation algorithms for wireless mesh networks, which are *optimal* with respect to resource utilization and *fair* across different network access points. Compared with traditional wireline networks, the unique characteristics of wireless mesh networks pose great challenges to such algorithms. Particularly, the wireless interference issue of mesh networks needs fresh treatment: flows not only contend at the same wireless mesh router (contention in the time domain), but also compete for the shared channel if they are within the interference ranges of each other (contention in the spatial domain). This challenge calls for a new resource allocation framework that could characterize the unique features of wireless mesh networks.

To address this challenge, we present a price-based resource allocation framework for wireless mesh networks to achieve *optimal* resource utilization and *fairness* among competing aggregated flows. In this chapter, we first model the resource allocation problem as an optimization problem: given network resources with constrained capacities and a set of users (*e.g.*, aggregated flows from access points of mesh networks), one tries to allocate resources to each user in a way that the overall satisfaction (so called *utility*) of all users are maximized. We show that such an optimization goal could naturally lead to different fairness objectives when appropriate utility functions are specified. We further present a price-based distributed algorithm which solves this optimization problem and thus provides *fair* and *optimal* resource allocation.

We instantiate the above generalized resource allocation framework to the wireless mesh networks. The key challenge comes from the *shared-medium multi-hop*

nature of such networks, namely location-dependent contention and spatial reuse. Based on solid theoretical analysis, we show that a resource element in a multihop wireless mesh network is a *facet of the polytope defined by the independent set of the conflict graph of this network, which could be approximated by a maximal clique*. Thus we build our price-based resource allocation framework on the notion of *maximal cliques* in wireless mesh networks, as compared to individual links in traditional wide-area wireline networks. We further present a price-based distributed algorithm, which is proven to converge to the global network optimum with respect to resource allocation. The algorithm is validated and evaluated through simulation study.

Our theoretical resource allocation framework of wireless mesh networks possesses great practical advantages. First, with the evolution of wireless signaling technology, medium access and routing protocols, the solution space of this problem may keep reforming, but its nature of optimal resource allocation remains unchanged. A good theoretical framework can effectively decouple the "core" of the problem and its other components (e.g., definition of network resource, and the way it is assigned to users), so that the basic problem formulation and its solution methodology survive. Second, perfect solutions often do not exist, since finding the optimal resource allocation (optimal point in the solution space) is always extremely expensive, if not impossible. When one designs practical solutions to approximate this optimal point, the role of a theoretical framework becomes crucial as it provides philosophical guidance of what is a good intuition.

The rest of this chapter is organized as follows. Section 6.2 introduces the generalized resource allocation framework. Section 6.3 instantiates this framework to the case of wireless mesh network. Section 6.4 presents the price-based decentralized resource allocation algorithm. Finally, we show simulation results in Section 6.5, discuss related works in Section 6.6, and then we conclude the chapter.

## 6.2 Theoretical Framework for Price-Based Resource Allocation

In this section, we present the generalized price-based theoretical framework for resource allocation in the setting of an abstract network model. We first formulate the resource allocation problem as an optimization problem. We then show that a price-based approach can provide a decentralized algorithm to solve this problem.

### 6.2.1 Resource Allocation: An Optimization Problem

#### An abstract network model

In our abstract network model, a network is represented as a set of *resource elements* $E$. A resource element $e \in E$ can be a wireline link, a shared wireless channel, etc. Each element has a fixed and finite capacity $C_e$. Note that the most important nature of a resource element is the independence of its capacity. Specifically, how resources are allocated can not affect the capacity of a resource element. In this sense, a wireline link is a resource element, while a wireless link is not, as its capacity may

vary depending on the traffic in its neighborhood and the scheduling algorithm in use. Characterizing the resource elements in a wireless mesh network is an important yet difficult issue, which will be elaborated in Section 6.3.

This network is shared by a set of flows (*e.g.*, end-to-end aggregated flows in mesh network) $F$. A flow $f \in F$ has a rate of $x_f$ and $f$ must traverse a sequence of resource elements (*i.e.*, the end-to-end path of $f$ passes multiple links) to reach its destination. Let $R_{ef}$ be the amount of resource $e$ used by a unit flow of $f$, and $y_e$ be the amount of traffic generated by all flows in $F$ through resource element $e$. Obviously $y_e = \sum_{f \in F} R_{ef} x_f$. Note that the calculation of $R_{ef}$ depends on the definition of resource element, which may vary for different types of networks.

**Objective: maximizing aggregated utility**

We associate each end-to-end flow $f \in F$ with a *utility function* $U_f(x_f) : \Re_+ \rightarrow \Re_+$, which represents the degree of satisfaction of its associated end user. Here we make the following assumptions about $U_f(x_f)$:

- **A1**. On the interval $[0, \infty)$, the utility function $U_f(\cdot)$ is increasing, strictly concave and continuously differentiable.
- **A2**. $U_f$ is additive so that the aggregated utility of rate allocation $\boldsymbol{x} = (x_f, f \in F)$ is $\sum_{f \in F} U_f(x_f)$.

We investigate the problem of optimal resource allocation in the sense of *maximizing the aggregated utility function* of all users, which is also referred to as the *social welfare* in the literature. Formally, this objective is given as follows,

$$\textbf{maximize} \sum_{f \in F} U_f(x_f).$$

This optimization objective is of particular interest. As we will demonstrate shortly, such an objective achieves Pareto optimality with respect to the resource utilization, and also realizes different fairness models — including proportional and max-min fairness — when appropriate utility functions are specified.

**Constraint: resource element and its capacity**

Recall that each element $e \in E$ in the network has a finite capacity $C_e$, and $y_e$ is the amount of traffic generated by all flows in $F$ through resource element $e$. The constraints on resource capacities are given as follows:

$$\forall e \in E, y_e \leq C_e.$$

Since $R_{ef}$ is the amount of resource $e$ used by a unit flow of $f$, we have $y_e = \sum_{f \in F} R_{ef} \cdot x_f$. Thus the resource constraint is given as follows:

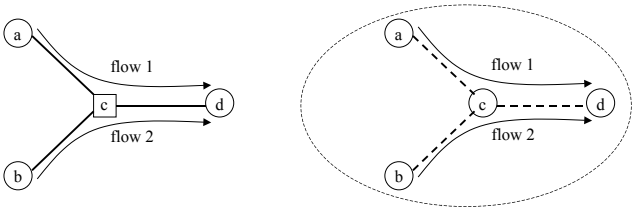
$$\forall e \in E, \sum_{f \in F} R_{ef} \cdot x_f \leq C_e$$

or in concise form,

$$\boldsymbol{R} \cdot \boldsymbol{x} \leq \boldsymbol{C}$$

where $\boldsymbol{R} = (R_{ef})_{|E|\times|F|}$ is a matrix with element $R_{ef}$ at row $e$ and column $f$, and $\boldsymbol{x} = (x_f, f \in F)$, $\boldsymbol{C} = (C_e, e \in E)$ are vectors of flow rates and resource capacities, respectively.

The definition of resource element and its capacity can vary for different types of networks. It is particularly hard to define for wireless mesh networks. In what follows, we will illustrate the concept of resource element in two simple network settings and present their resource constraints. Based on these intuitive examples, we will further define the resource model of a wireless mesh network in Section 6.3, which establishes the foundation of theoretical study on resource allocation in this type of network.




(a) Wireline network (b) Wireless network without spatial reuse

**Fig. 6.1.** Resource elements in different networks.

*Wireline networks*

In wireline networks, flows only contend with each other if they share the same physical link. In this case, the resource element $e$ is a wireline link, its resource capacity $C_e$ is the link capacity. In this case, $\boldsymbol{R}$ can be understood as the routing matrix defined as follows.

$$R_{ef} = \begin{cases} 1, & \text{if } f \text{passes through } e \\ 0, & \text{otherwise.} \end{cases}$$

In the example shown in Fig. 6.1 (a), the constraints on resource allocations of flows 1 and 2 can be expressed as

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} C_{ac} \\ C_{bc} \\ C_{cd} \end{pmatrix}.$$

*Wireless networks without spatial reuse*

We now consider a simple wireless network based on unit disk graph model as shown in Fig. 6.1 (b). All four nodes are within the transmission range of each other

and have the same data transmission rate. Flow 1 and 2 not only contend at the wireless link $\{c, d\}$ which they both traverse, but also at the link $\{a, c\}$ and $\{b, c\}$ which share the same wireless channel. Hence in this case, the wireless channel shared by these three links is the only resource element, whose capacity is $C_{chan}$ – the wireless channel capacity. Since each flow passes two hops in this wireless channel, $R_{ef_1} = R_{ef_2} = 2$. Then the constraint on resource allocation of flow 1 and 2 can be expressed as

$$\begin{pmatrix} 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq C_{chan}.$$

**Putting things together**

Summarizing the above discussions, we formulate the resource allocation problem in a generalized form as follows:

$$\mathbf{S}: \textbf{maximize} \sum_{f \in F} U_f(x_f) \tag{6.1}$$

$$\textbf{subject to } \boldsymbol{R} \cdot \boldsymbol{x} \leq \boldsymbol{C} \tag{6.2}$$

$$\boldsymbol{x} \geq \mathbf{0}. \tag{6.3}$$

The objective function in (6.1) maximizes the aggregated utility of all flows. The constraint of the optimization problem (inequality (6.2)) comes from the resource constraint of the network. We now demonstrate that, by optimizing towards such an objective, both *optimal resource utilization* and *fair resource allocation* may be achieved among *end-to-end* flows. *Pareto optimality* With respect to optimal resource

utilization, we show that the resource allocation is *Pareto optimal* if the optimization problem $\mathbf{S}$ can be solved. Formally, Pareto optimality is defined as follows.

**Definition 1. (Pareto optimality)** *A rate allocation* $\boldsymbol{x} = (x_f, f \in F)$ *is* **Pareto optimal**, *if it satisfies the following two conditions: (1)* $\boldsymbol{x}$ *is feasible,* i.e., $\boldsymbol{x} \geq \mathbf{0}$ *and* $\boldsymbol{R} \cdot \boldsymbol{x} \leq \boldsymbol{C}$*; and (2)* $\forall \boldsymbol{x}'$ *which satisfies* $\boldsymbol{x}' \geq \mathbf{0}$ *and* $\boldsymbol{R} \cdot \boldsymbol{x}' \leq \boldsymbol{C}$*, if* $\boldsymbol{x}' \geq \boldsymbol{x}$*, then* $\boldsymbol{x}' = \boldsymbol{x}$*. In the second condition, the* $\geq$ *relation is defined such that, two vectors* $\boldsymbol{x}$ *and* $\boldsymbol{x}'$ *satisfy* $\boldsymbol{x}' \geq \boldsymbol{x}$*, if and only if for all* $f \in F$*,* $x'_f \geq x_f$*.*

**Proposition 1.** *A rate allocation* $\boldsymbol{x}$ *is Pareto optimal, if it solves the problem* $\mathbf{S}$*, with increasing utility functions* $U_f(x_f)$*, for* $f \in F$*.*

*Proof.* Let $\boldsymbol{x}$ be a solution to the problem $\mathbf{S}$. If $\boldsymbol{x}$ is not Pareto optimal, then there exists another vector $\boldsymbol{x}' \neq \boldsymbol{x}$, which satisfies $\boldsymbol{R} \cdot \boldsymbol{x}' \leq \boldsymbol{C}$ and $\boldsymbol{x}' > \boldsymbol{x}$. As $U_f(\cdot)$ is increasing, we have $\sum_{f \in F} U_f(x'_f) > \sum_{f \in F} U_f(x_f)$. This leads to a contradiction, as $\boldsymbol{x}$ is the solution to $\mathbf{S}$ and hence maximizes $\sum_{f \in F} U_f(x_f)$.

*Fairness* By choosing appropriate utility functions, the optimal resource allocation can implement different fairness models among the flows. We illustrate this fact using two commonly adopted fairness models: *weighted proportional* fairness and *max-min* fairness.

**Definition 2. (weighted proportional fairness)** *A vector of rates* $\boldsymbol{x} = (x_f, f \in F)$ *is* **weighted proportionally fair** *with the vector of weights* $w_f$*, if it satisfies the following two conditions: (1)* $\boldsymbol{x}$ *is feasible,* i.e.*,* $\boldsymbol{x} \geq 0$ *and* $\boldsymbol{R} \cdot \boldsymbol{x} \leq \boldsymbol{C}$*; and (2) for any other feasible vector* $\boldsymbol{x}' = (x'_f, f \in F)$*, the aggregation of proportional changes is zero or negative:*

$$\sum_{f \in F} w_f \frac{x'_f - x_f}{x_f} \leq 0.$$

**Proposition 2.** *A rate allocation* $\boldsymbol{x}$ *is weighted proportional fair with the weight vector* $w_f$*, if and only if it solves the problem* $\mathbf{S}$*, with* $U_f(x_f) = w_f \log x_f$ *for* $f \in F$*.*

*Proof.* As shown in [8], by the optimality condition (6.1), this proposition can be derived according to the following relation:

$$\sum_{f \in F} \frac{\partial U_f}{\partial x'_f}(x_f)(x'_f - x_f) = \sum_{f \in F} w_f \frac{x'_f - x_f}{x_f} < 0$$

where the strict inequality follows from the strict concavity of $U_f$.

**Definition 3. (max-min fairness)** *A vector of rates* $\boldsymbol{x} = (x_f, f \in F)$ *is* **max-min fair***, if it satisfies the following two conditions: (1)* $\boldsymbol{x}$ *is feasible,* i.e.*,* $\boldsymbol{x} \geq 0$ *and* $\boldsymbol{R} \cdot \boldsymbol{x} \leq \boldsymbol{C}$*; and (2) for any* $f \in F$*, increasing* $x_f$ *can not be achieved without decreasing the fair share* $x_{f'}$ *of another flow* $f' \in F$ *that satisfies* $x_f \geq x_{f'}$*.*

**Proposition 3.** *A rate allocation* $\boldsymbol{x}$ *is max-min fair if and only if it solves the problem* $\mathbf{S}$*, with* $U_f(x_f) = -(-\log x_f)^{\zeta}$*,* $\zeta \to \infty$ *for* $f \in F$*.*

These results straightforwardly follow their counterparts in wireline networks [8].

### 6.2.2 Decentralized Solution: A Price-Based Approach

We proceed to study the decentralized solution to the problem $\mathbf{S}$ so that the optimal resource allocation can be achieved.

By assumption **A1**, the objective function of $\mathbf{S}$ in (6.1) is differentiable and strictly concave. In addition, the feasible region of the optimization problem in inequality (6.2) is convex and compact. By non-linear optimization theory, there exists a unique maximizing value of argument $\boldsymbol{x}$ for the above optimization problem. Let us consider the Lagrangian form of the optimization problem $\mathbf{S}$:

$$L(\boldsymbol{x};\boldsymbol{\mu}) = \sum_{f\in F} U_f(x_f) + \boldsymbol{\mu}^T(\boldsymbol{C} - \boldsymbol{R}\boldsymbol{x}) \tag{6.4}$$
$$= \sum_{f\in F}(U_f(x_f) - x_f \sum_{e\in E} \mu_e R_{ef}) + \sum_{e\in E} \mu_e C_e$$

where $\boldsymbol{\mu} = (\mu_e, e \in E)$ is a vector of Lagrange multipliers. Given global knowledge of utility functions, $\mathbf{S}$ is mathematically tractable. However, in practice, such knowledge is unlikely to be available. In addition, it may be infeasible to compute and allocate resources in a centralized fashion. Here we seek a decentralized solution. The key to decentralization is pricing.

In the Lagrangian form specified in (6.4), the Lagrange multipliers $\mu_e$ may be regarded as the implied cost, or the *shadow price*, of a unit flow using resource $e$. Such a price $\mu_e$ reflects the traffic load $y_e$ at the resource element $e$. Flow $f \in F$ will then be charged with a flow price $\lambda_f$ which is the sum of the costs of all resource elements it uses, the cost of each resource element being the product of its price and the amount of resource used by a unit flow of $f$, namely,

$$\lambda_f = \sum_{e\in E} R_{ef}\mu_e.$$

Based on the flow price $\lambda_f$, flow $f$ can make a self-optimized decision to adjust its sending rate $x_f$. The aggregated sending rate $y_e = \sum_{f\in F} R_{ef} \cdot x_f$ of all flows in resource element $e$ in turn affects its price $\mu_e$. To summarize, Fig. 6.2 illustrates the price-based resource allocation framework. Here we deem each component in the diagram as abstract entities capable of computing and communicating. This framework involves no central authority and purely depends on local decision of each component and exchange of control signals among them. In each cycle, a resource element $e$ calculates its load $y_e$, the total amount of flows passing through it, then derives its penalty $\mu_e$ and sends it to all these flows. Meanwhile, a flow $f$, on receiving prices from all resource elements it traverses, derives its flow price $\lambda_f$, then adjusts flow rate $x_f$. Such a cycle repeats itself, and finally converges to an equilibrium point.

The presented framework is a generalized form of the framework proposed by Kelly et al. in [8, 9] for wireline networks. As we will show in Section 6.4, such a generalization is critical for us to study the resource allocation problem in wireless mesh networks. We list all notations introduced in Section 6.2 as follows.

## 6.3 Resource Model of Multihop Wireless Mesh Networks

In this section, we study the resource model and identify the resource elements of a wireless mesh network. We consider a two-tier wireless mesh network shown in Fig. 6.3. In this network, each end host accesses a local access point (LAP). These local access points, along with multiple stationary wireless routers, are also called

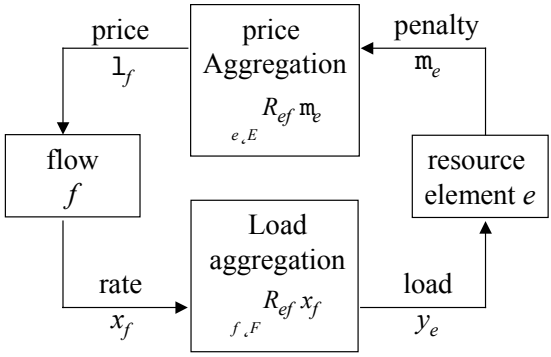



**Fig. 6.2.** Price-based resource allocation framework.

**Table 6.1.** Notations in Section 6.2.

| Notation | Definition |
|---|---|
| $f \in F$ | End-to-end flow in the network |
| $e \in E$ | Resource element of the network |
| $\boldsymbol{x} = (x_f, f \in F)$ | Rate vector of flow $f \in F$ |
| $\boldsymbol{C} = (C_e, e \in E)$ | Capacity vector of resource element $e \in E$ |
| $\boldsymbol{R} = (R_{ef})_{\|E\|\times\|F\|}$ | Resource constraint matrix |
| $U_f(x_f)$ $(f \in F)$ | Utility function of flow $f \in F$ |
| $\boldsymbol{y} = (y_e, e \in E)$ | Aggregated traffic load at resource element $e \in E$ |
| $\boldsymbol{\mu} = (\mu_e, e \in E)$ | Price of resource element $e \in E$ |
| $\lambda_f$ $(f \in F)$ | Price of flow $f \in F$ |

mesh nodes. These nodes communicate with each other and form a multi-hop wireless backbone. This backbone network eventually forwards user traffic to the gateway access points (GAPs) connected to the Internet via physical wireline connection.

This chapter focuses on the resource allocation issue in wireless mesh backbone network. The goal is to achieve fairness among local access points. In particular, we consider a wireless mesh *backbone* network that consists of a set of nodes $N$. The transmission of each node $n_i \in N$ follows the unit disk graph model with a transmission range of $d_{tx}$ and an interference range of $d_{int}$, which can be larger than $d_{tx}$.

To simplify the discussion, we only consider the scenario where mesh nodes use the same wireless channel. Packet transmission in such a network is subject to location-dependent contention. Here we consider the protocol model proposed in [10]. In this model, the transmission from node $n_i$ to $n_j$ $(n_i, n_j \in N)$ is successful if (1) the distance between these two nodes $d_{ij}$ satisfies $d_{ij} < d_{tx}$, and (2) any node $n_k \in N$, which is within the interference range of the receiving node $n_j$ $(d_{kj} \leq d_{int})$, is not transmitting. This model can be further refined to the case of IEEE 802.11-style MAC protocol, where the sending node $n_i$ is also required to be free of interference as it needs to receive the link layer acknowledgement from the receiving

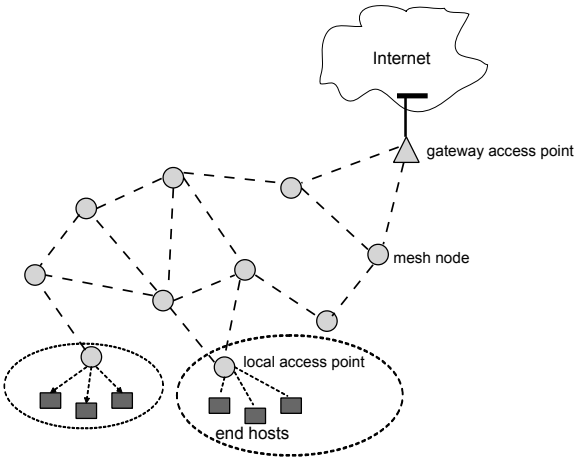



**Fig. 6.3.** An example of wireless mesh network.

node $n_j$. Specifically, any node $n_k \in N$, which is within the interference range of $n_i$ or $n_j$ ($d_{kj} \leq d_{int}$ or $d_{ki} \leq d_{int}$), is not transmitting. We model such a network as a directional graph $G = (N, L)$, where $L \subseteq N^2$ denotes the set of wireless links.

Now let us consider a *conflict graph* $G_c = (V_c, L_c)$ of network $G$ [11]. A vertex of the conflict graph $v_i \in V_c$ corresponds to a wireless link in the network $l \in L$. There exists an edge between two vertices if the transmissions along these two wireless links contend with each other according to the above protocol model.

To illustrate these concepts, we show an example in Fig. 6.4. Fig. 6.4 (a) gives the network topology and the traffic used in the example. In this example, the transmission and interference range of a node is 250m and 550m, respectively; $a$ and $b$, $c$ and $d$, $e$ and $f$ are 250m apart; $b$ and $c$, $d$ and $e$ are 300m apart. Thus the wireless links $\{a, b\}$ and $\{c, d\}$ contend with each other, also do $\{c, d\}$ and $\{e, f\}$. But $\{a, b\}$ and $\{e, f\}$ can transmit simultaneously. The conflict graph of this wireless network is shown in Fig. 6.4 (b).

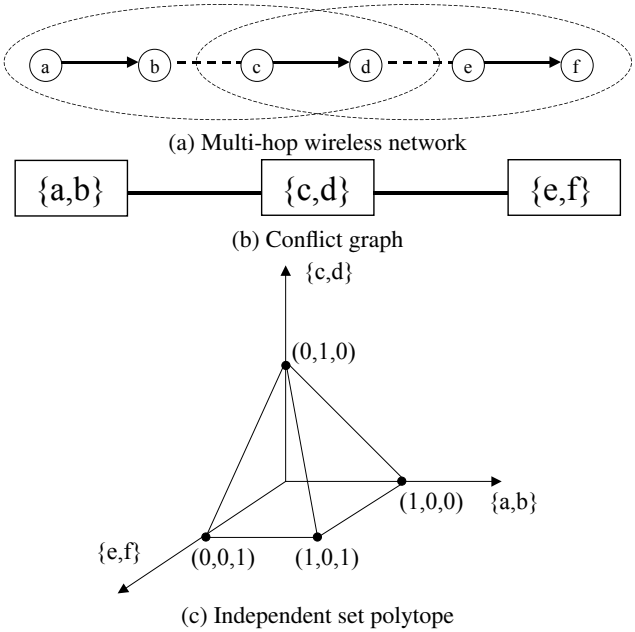
### 6.3.1 Identifying Resource Elements

Now let us consider an *independent set* $I \subseteq V_c$ of the graph $G_c$. $I$ can be represented using a $|V_c|$-dimension *independence vector* $\boldsymbol{\iota}_I = (\iota_j, v_j \in V_c)$, defined as follows:

$$\iota_j = \begin{cases} 1, & \text{if } v_j \in I \\ 0, & \text{otherwise} \end{cases}$$

where $\boldsymbol{\iota}_I$ can be regarded as a point in a $V_c$-dimensional *independence space*. In this space, each dimension corresponds to a vertex $v_i \in V_c$.

In the above example, besides the independent sets consisting of each vertex itself, $\{\{a, b\}, \{e, f\}\}$ is also an independent set. Let vertices $\{a, b\}$, $\{c, d\}$, $\{e, f\}$




(a) Multi-hop wireless network

(b) Conflict graph

(c) Independent set polytope

**Fig. 6.4.** Resource model of multi-hop wireless network.

correspond to the three dimensions of the independence space, the following independence vectors are shown in Fig. 6.4 (c): $(1,0,0)$,$(0,1,0)$,$(0,0,1)$ and $(1,0,1)$. A special independence vector is the origin point $(0,0,0)$.

The picture also shows that a polytope is formed as the convex hull of all points corresponding to each independence vector, or in other words, convex combination of all independence vectors. We call such a polytope the *independent set polytope*, denoted as $T_G$. Let us consider a $|V_c|$-dimension vector $\boldsymbol{q} = (q_j, v_j \in V_c)$, where $q_j$ is the fraction of time during which link $l$ corresponding to $v_j$ is active. Vector $\boldsymbol{q}$ is *schedulable* if there exists a collision-free MAC transmission schedule that allocates $q_j$ to link $l$ which corresponds to $v_j$. The result of [11] shows that

**Proposition 4**. *Vector $\boldsymbol{q} = (q_j, v_j \in V_c)$ is schedulable if and only if it lies within the independent set polytope $T_G$.*

Reflected in Fig. 6.4 (c), all points within the polytope $T_G$ is schedulable. To model the resource element from this concept, we consider the *facets* of the polytope $T_G$[3]. Note that we can get these facets by running any polynomial-time convex hull algorithm [12] on all vertices of the polytope (independence vectors). We collect all facets into a set $\boldsymbol{\Phi}$. The plane that a facet $\boldsymbol{\phi}_i \in \boldsymbol{\Phi}$ belongs to can be presented in the following linear form:

$$\sum_{j=1}^{|V_c|} \phi_{ij} q_j - Z_i = 0$$

[3]The facets that lie along the coordination plane are excluded.

where $\phi_{ij}$ and $Z_i$ are coefficients of the plane function. If we formulate $\boldsymbol{\Phi}$ into a matrix $\boldsymbol{\Phi} = (\phi_{ij})_{|\boldsymbol{\Phi}|\times|V_c|}$, then the polytope $T_G$ can be represented in the following vector form.

$$\boldsymbol{\Phi} \cdot \boldsymbol{q} \leq \boldsymbol{Z}$$

where $\boldsymbol{Z} = (Z_i, \boldsymbol{\phi}_i \in \boldsymbol{\Phi})$.

When the data rates of all wireless links are the same, we call such a rate the capacity of the wireless channel and denote it as $C_{chan}$. If we scale the independent set polytope $T_G$ by $C_{chan}$, then under the ideal centralized MAC scheduling, this polytope represents the solution space of our problem. In other words, a *wireless link rate allocation* $\boldsymbol{y} = (y_l, l \in L)$ is feasible, if the following condition holds:

$$\boldsymbol{\Phi} \cdot \boldsymbol{y} \leq C_{chan} \cdot \boldsymbol{Z}. \tag{6.5}$$

Let $\boldsymbol{A} = (A_{lf})_{|L|\times|F|}$ be the routing matrix defined as follows:

$$A_{lf} = \begin{cases} 1, & \text{if flow } f \text{ passes through wireless link } l \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that $\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{y}$. Substituting inequality (6.5) into the constraint of problem **S** in (6.2), we can easily derive that $\boldsymbol{R} = \boldsymbol{\Phi} \cdot \boldsymbol{A}$ and $\boldsymbol{C} = C_{chan} \cdot \boldsymbol{Z}$. From these properties, we observe that *each facet $\boldsymbol{\phi}_i \in \boldsymbol{\Phi}$ can be regarded as a resource element with an independent capacity $C_{chan} \cdot Z_i$*. Though the concept of independent set has been used in the existing works to explore the throughput limit of multihop wireless networks [14], the concept of facets of independent set polytope is never discussed. However, constructing the solution space by the formulation of facets is critical to the development of a decentralized algorithm.

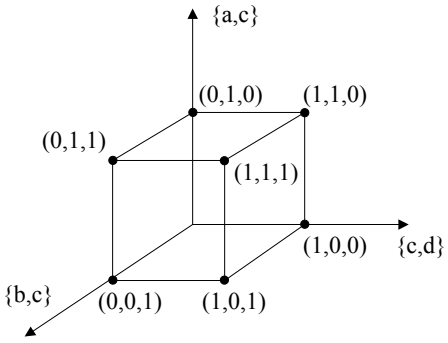



**Fig. 6.5.** Solution space of wireline network in Fig. 6.1 (a).

As an interesting finding, we also notice that the wireline network model is a special case of this formulation. Take the network shown in Fig. 6.4 (a) as an example. Since all links are independent from each other, any subset of the link set $L$ is an independent set. The resulting independent set polytope (with each dimension normalized by the capacity of its corresponding link) is shown in Fig. 6.5. An important

property of this polytope is that it is a *cube*. That means for each of its facets, the plane it belongs to only intersects with one axis. Since each axis represents a wireline link, back to Inequality (6.5), this implies that (1) $\boldsymbol{\Phi} = \boldsymbol{I}$, the identity matrix, thus $\boldsymbol{A} = \boldsymbol{R}$, and (2) $Z_i = 1$ for any $\boldsymbol{\phi}_i \in \boldsymbol{\Phi}$.

Note that the above formulation could be easily extended to the case of heterogeneous wireless link data rates. We denote the wireless link data rates using a vector $\boldsymbol{b} = (b_l, l \in L)$. It is obvious that $q_j = \frac{y_l}{b_l}$ for $v_j$ which corresponds to $l$. Let $\boldsymbol{b}' = (1/b_l, l \in L)$. It is obvious that $\boldsymbol{q} = \boldsymbol{y} \cdot \boldsymbol{b}'^T$. Thus the constraint for wireless link rate allocation is given as follows:

$$\boldsymbol{\Phi} \cdot \boldsymbol{y} \cdot \boldsymbol{b}'^T \leq \boldsymbol{Z}.$$

For simplicity, we only consider the homogeneous wireless link rate in the following discussions.

### 6.3.2 Approximating Resource Element

To this end, we have clearly identified the resource elements of a multihop wireless mesh backbone network. However, applying this model to the resource allocation framework can still be difficult, as the problem of finding all independent sets is NP-hard. Besides, this model assumes ideal MAC scheduling. It is difficult to be applied for practical implementation in realistic wireless network settings, (e.g., with non-ideal MAC algorithm such as IEEE 802.11), because the facets of the independent set polytope of the contention graph lack the intuition to be mapped into any instance in the physical wireless network, not to mention to be implemented via distributed algorithms.

To address this difficulty, we explore the approximation of resource elements by studying the upper bound of the resource constraint in a multihop wireless network. Here we present a *maximal-clique-based* approximation. Such an approximation gives a good intuitive explanation on the structure of the resource element in the physical network. Thus it can also facilitate the distributed implementation of resource allocation algorithms.

In a graph, a complete subgraph is referred to as a *clique*. A *maximal clique* is defined as a clique that is not contained in any other cliques[4]. In a contention graph, the vertices in a maximal clique represent a maximal set of mutually contending wireless links, along which at most one subflow may transmit at any given time. Intuitively, each maximal clique in a contention graph represents a maximal distinct contention region, since at most one subflow in the clique can transmit at any time, and adding any other flows into this clique will introduce the possibility of simultaneous transmissions. We denote the set of all maximal cliques in $G_c$ as $Q$.

[4]Note that the *maximal clique* has a different definition from the *maximum clique* of a graph, which is the maximal clique with the largest number of vertices. Finding the maximum clique of a graph is a NP-complete problem, while enumerating all the maximal cliques of a graph can be solved in polynomial time [13].

Based on the above discussions, we have the following results for rate allocation vector $\boldsymbol{y}$:

**Proposition 4.** *If rate allocation* $\boldsymbol{y} = (y_l, l \in L)$ *is feasible, then the following condition is satisfied.*

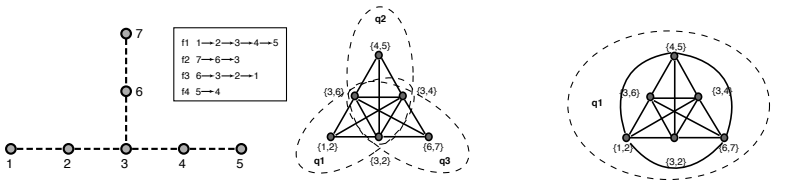
$$\forall e \in Q, \sum_{l \in V(e)} y_l \leq C_{chan} \tag{6.6}$$

*where* $V(e) \subseteq L$ *is the set of vertices in clique* $e$*.*

Eq. (6.6) gives an upper bound on the rate allocations to the wireless links. Such a bound may not be tight. First, there may exist no schedules that assign rates to the wireless links to achieve this bound. Such scenario happens when the contention graph has odd holes or odd anti-holes [14]. Second, for some contention graphs, even if there exists an ideal centralized scheduling algorithm that can achieve this bound, the distributed scheduling algorithms that employ carrier sensing multiple access (*e.g.*, IEEE 802.11) can not achieve this bound. To address the above issues, we introduce $C_e$, the *achievable* channel capacity at a clique $e$ so that if $\sum_{l \in V(e)} y_l \leq C_e$ then $\boldsymbol{y} = (y_l, l \in L)$ is feasible. To this end, we observe that a maximal clique $e$ can be regarded as an approximation of a resource element with capacity $C_e$.

We now proceed to consider the resource constraint of a wireless mesh backbone network using the maximal clique as the approximation of the resource element. In particular, we define a clique-flow matrix $\boldsymbol{R} = \{R_{ef}\}$, where $R_{ef} = |V(e) \cap L(f)|$ represents the number of subflows that flow $f$ has in the clique $e$. If we treat a maximal clique as a resource element, then the clique-flow matrix $\boldsymbol{R}$ represents the "resource usage pattern" of each flow. Let the vector $\boldsymbol{C} = (C_e,\ e \in Q)$ be the vector of achievable channel capacities in each of the cliques. Constraints with respect to rate allocations to end-to-end aggregated flows are presented in the following proposition.

**Proposition 5.** *In a multi-hop wireless network* $G = (N, L)$ *with a set of flows* $F$*, there exists a feasible rate allocation* $\boldsymbol{x} = (x_f, f \in F)$*, if and only if* $\boldsymbol{R} \cdot \boldsymbol{x} \leq \boldsymbol{C}$*.*




(a) Network topology (b) Wireless link contention graph ($d_{tx} = d_{int}$) (c) Wireless link contention graph ($2d_{tx} = d_{int}$)

**Fig. 6.6.** Approximate resource model of multihop wireless network.

We present an example to illustrate the above concepts and notations. Fig. 6.6(a) shows the topology of the network, as well as its ongoing flows. The corresponding contention graph is shown in Fig. 6.6(b). In this example, there are $4$ end-to-end flows $f_1 = \{\{1,2\},\{2,3\},\{3,4\},\{4,5\}\}$, $f_2 = \{\{7,6\},\{6,3\}\}$, $f_3 =$

$\{\{6,3\},\{3,2\},\{2,1\}\}$ and $f_4 = \{\{5,4\}\}$. As such, in Fig. 6.6(b) there are three maximal cliques in the contention graph: $e_1 = \{\{1,2\},\{3,2\},\{3,4\},\{3,6\}\}, e_2 = \{\{3,2\},\{3,4\},\{4,5\},\{3,6\}\}$ and $e_3 = \{\{3,2\},\{3,4\},\{3,6\},\{6,7\}\}$.

We use $y_{ij}$ to denote the aggregated rate of *all* subflows along wireless link $\{i,j\}$. For example, $y_{12} = x_1 + x_3$, $y_{36} = x_2 + x_3$. In each clique, the aggregated rate may not exceed the corresponding channel capacity. That is

$$y_{12} + y_{32} + y_{34} + y_{36} \leq C_1 \tag{6.7}$$

$$y_{32} + y_{34} + y_{45} + y_{36} \leq C_2 \tag{6.8}$$

$$y_{32} + y_{34} + y_{36} + y_{67} \leq C_3. \tag{6.9}$$

When it comes to end-to-end flow rate allocation, the resource constraint imposed by shared wireless channels is as follows:

$$\begin{pmatrix} 3\ 1\ 3\ 0 \\ 3\ 1\ 2\ 1 \\ 2\ 2\ 2\ 0 \end{pmatrix} \cdot \boldsymbol{x} \leq \boldsymbol{C}.$$

We collect the notations introduced in this section into Table 6.2.

**Table 6.2.** Notations in Section 6.3.

| Notation | Definition |
|---|---|
| $n_i \in N$ | Wireless node |
| $l = \{n_i, n_j\} \in L$ | Wireless link connecting nodes $n_i$ and $n_j$ |
| $d_{ij}$ | Distance between nodes $n_i$ and $n_j$ |
| $G = (N, L)$ | Wireless network |
| $G_c = (N_c = L, L_c)$ | Conflict graph of $G$ |
| $I \subseteq V_c$ | Independent set of $G_c$ |
| $\boldsymbol{\iota}_I = (\iota_j, v_j \in V_c)$ | Independence vector of $I$ |
| $T_G$ | Independent set polytope of $G_c$ |
| $\boldsymbol{q} = (q_j, v_j \in V_c)$ | Active time scheduling vector for all links |
| $\boldsymbol{y} = (y_l, l \in L)$ | Link flow scheduling vector |
| $\boldsymbol{\Phi} = (\phi_{ij})_{\lvert\boldsymbol{\Phi}\rvert \times \lvert V_c \rvert}$ | Facet matrix of polytope $T_G$ |
| $\boldsymbol{Z} = (Z_i, \boldsymbol{\phi}_i \in \boldsymbol{\Phi})$ | Facet coefficient vector of polytope $T_G$ |
| $C_{chan}$ | Wireless channel capacity |
| $\boldsymbol{A} = (R_{lf})_{\lvert L \rvert \times \lvert F \rvert}$ | Routing matrix |

## 6.4 Price-Based Resource Allocation Algorithm

We now present the decentralized algorithms for resource allocation in multihop wireless networks based on the theoretical framework in Section 6.2 and the resource model in Section 6.3.

### 6.4.1 Price Model

We first illustrate the concepts and components of the generalized resource allocation framework in the setting of wireless mesh network. Recall that a resource element $e$ in multihop wireless networks is a set of wireless links defined by a facet of independent set polytope $\phi_i \in \boldsymbol{\Phi}$, which could be approximated by a maximal clique $e \in Q$. Thus the amount of traffic $y_e$ at the resource element $e$ is the sum of traffic at the wireless links that belong to the resource element $e$:

$$y_e = \sum_{l \in e} y_l.$$

As a wireless link $l$ may be the member of several resource elements $e$, we define the price of wireless link as follows:

$$\mu_l = \sum_{e: l \in e} \mu_e.$$

Thus the price of a flow $f$ can be represented in following two alternative ways:

$$\lambda_f = \sum_{e \in E} R_{ef} \mu_e \tag{6.10}$$

$$= \sum_{l: f \text{ passes } l} \mu_l. \tag{6.11}$$

The first representation (6.10) can be explained as follows. Flow $f$ needs to pay for all the resource elements it uses. For each resource element, the cost is the product of the number of wireless links that $f$ traverses in this resource element and its price. In the second representation (6.11), flow price is the aggregated price of all wireless links it passes. Note that for each wireless link, its price is the aggregated price of all the resource elements that it belongs to.

### 6.4.2 Price-Based Rate Limiting Algorithm

Assume that resource element prices $\boldsymbol{\mu} = (\mu_e, e \in E)$ are generated appropriately as a function of the load $\boldsymbol{y} = (y_e, e \in E)$ at these resource elements. We first study how flows adjust their resource usages.

As presented in the theoretical framework in Section 6.2, flow $f$ attempts to maximize its net benefit.

$$\max_{x_f} \{U_f(x_f) - \lambda_f \cdot x_f\}.$$

A simple first-order condition establishes that

$$U_f'(x_f) = \lambda_f.$$

Thus it adapts its rates to equalize the flow price, *i.e.*, $\lambda_f = \sum_{e \in E} \mu_e R_{ef}$, with a target value $U'_f(x_f)$. Formally, the rate adaptation algorithm of flow $f$ can be represented in the following differential equation:

$$\frac{d}{dt}x_f(t) = \gamma\left(1 - \frac{1}{U'_f(x_f(t))}\sum_{e \in E}\mu_e(t)R_{ef}\right) \tag{6.12}$$

where $x_f(t)$ is the rate of $f$ at time $t$. The price $\mu_e(t) = \mu_e(y_e(t))$ is a non-negative, continuous and increasing function of the total traffic $y_e = \sum_{f \in F} R_{ef}x_f$ at resource $e$ at time $t$, and $\gamma$ is the amount of adjustment. Alternatively, it can be represented in the discrete time form:

$$x_f[t+1] = x_f[t] + \gamma\left(1 - \frac{1}{U'_f(x_f[t])}\sum_{e \in E}\mu_e[t]R_{ef}\right). \tag{6.13}$$

We now establish the stability of this algorithm. Further we show that, at equilibrium each flow maximizes its own net benefit; moreover, the flows collectively solve the relaxation of the original problem. This result is formally presented in the following theorem.

**Theorem 1**. *Let*

$$\mathcal{V}(\boldsymbol{x}) = \sum_{f \in F} U_f(x_f) - \sum_{e \in E}\int_0^{y_e} \mu_e(z)R_{ef}dz.$$

*$\mathcal{V}(\boldsymbol{x})$ is a strictly concave function. Moreover, it is a Lyapunov function for the system of the differential equation (6.12). The unique value $\boldsymbol{x}$ that maximizes $\mathcal{V}(\boldsymbol{x})$ is a stable point of the system, to which all trajectories converge. In addition, $\boldsymbol{x}$ is the unique equilibrium of the discrete time system specified by (6.13).*

*Proof.* Observe that

$$\frac{\partial \mathcal{V}(\boldsymbol{x})}{\partial x_f} = U'_f(x_f) - \sum_{e \in E}\mu_e(y_e)R_{ef}. \tag{6.14}$$

Setting these derivatives to zero identifies the maximum. Further

$$\frac{d\mathcal{V}(\boldsymbol{x}(t))}{dt} = \sum_{f \in F}\frac{\partial \mathcal{V}(\boldsymbol{x})}{\partial x_f}\cdot\frac{dx_f(t)}{dt} \tag{6.15}$$

$$= \gamma\sum_{f \in F}U'_f(x_f)\left(1 - \frac{1}{U'_f(x_f)}\mu_e(y_e)R_{ef}\right)^2 \tag{6.16}$$

establishes that $\mathcal{V}(\boldsymbol{x}(t))$ is strictly increasing with $t$, unless $\boldsymbol{x}(t) = \boldsymbol{x}^*$, the unique $\boldsymbol{x}$ maximizing $\mathcal{V}(\boldsymbol{x})$. The function $\mathcal{V}(x)$ is thus a Lyapunov function for the system (6.12), which establishes the result of Theorem 1.

The presented rate adaptation algorithm can be implemented as an ingress rate limiting mechanism for aggregated flows at local access points as in [15].

### 6.4.3 Discussion

The clique-based resource element definition is based on the assumption of ideal MAC scheduling. Such a definition helps us to define the upper bound of the solution space of the resource allocation problem. Yet, in practice, MAC algorithms, such as IEEE 802.11, can perform much worse than the ideal one. Also in order to calculate the price of a clique, the mesh routers need to communicate with each other to exchange their load, which may incur unnecessary overhead [16]. In recognition of the hardness of this problem, here we present a heuristic algorithm for price generation in IEEE 802.11-based networks. The main propose of this algorithm is to illustrate how the presented theoretical framework can be applied to real network settings.

Our approximation is based on two observations. First, due to the characteristics of conflict graph, the wireless links within a clique are most likely to be close to each other geographically. Second, contention window size in IEEE 802.11 can give important hints for traffic load in the neighborhood. Based on these observations, we make the following approximation in the calculation. First, we consider a resource element $e$ consisting of a wireless link $l$ and its neighborhood wireless links $l'$ which has one node that is within the range of virtual carrier sense, *i.e.*, one node of $l'$ can hear the RTS or CTS sent from wireless link $l$. Second, we use the contention window sizes $cw$ of the nodes connecting these links to infer the traffic at this resource element. Formally, the price $\mu_e$ of resource element $e$ is generated as follows:

$$\mu_e(t) = \beta \cdot m(\bar{cw}_e(t), \tilde{cw}). \tag{6.17}$$

In (6.17), $\beta$ is a scaling factor. $\bar{cw}_e(t)$ is the average contention window size within resource element $e$ at time $t$. Nodes exchange the information of their contention window sizes via piggybacking them onto RTS/CTS control frames, $\tilde{cw}$ is the target contention window size, which is a tunable parameter of this implementation. An ideal target contention window size needs to be tuned according the node density of the network. Function $m(\bar{cw}_e(t), \tilde{cw})$ is defined as the probability that $\bar{cw}_e(t)$ is larger than $\tilde{cw}$. It is easy to see that $m(\bar{cw}_e(t), \tilde{cw})$ is an increasing function of $\bar{cw}_e$ and a decreasing function of $\tilde{cw}$.

## 6.5 Performance Evaluation

In this section, we evaluate the performance of our resource allocation algorithm. We implement the algorithm based on the wireless extensions in ns-2. In the simulation, the physical wireless channel capacity is 1 Mbps; the utility function is $U_f(x_f) = \log(x_f)$; the packet size is 1000 bytes; the transmission range and interference range are $250m$ and $550m$, respectively. We use static shortest path routing as the routing protocol. The algorithm is simulated over two simple mesh network topologies as shown in Fig. 6.7. We study the performance of resource allocation algorithm in terms of system stability and fairness for flows with different lengths.

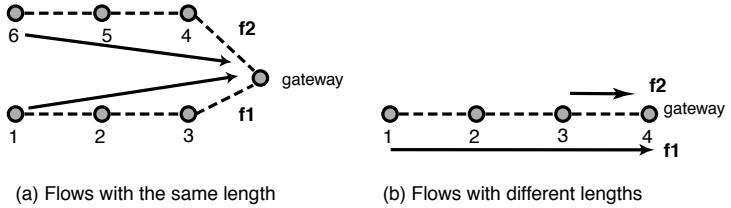



**Fig. 6.7.** Simulation topologies.

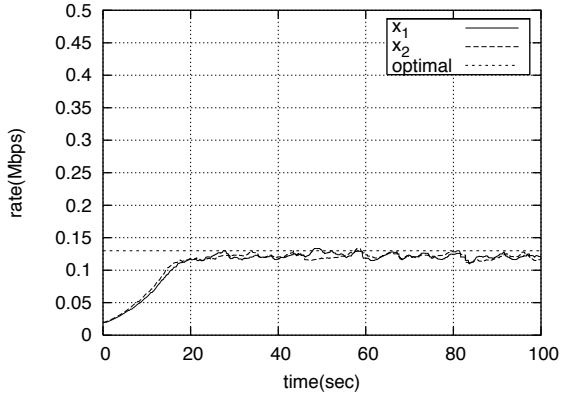
We first study the instantaneous behavior of the resource allocation algorithm and investigate the stability of the system with different settings of parameter $\gamma$ and initial flow rate $x_f$. In this experiment, our algorithm is simulated over the topology in Fig. 6.7(a). The default parameter values are set as follows: $\gamma = 0.005$, $\beta = 10$, and $x_1(0) = x_2(0) = 200$ Kbps. Fig. 6.8 plots the instantaneous throughput of the system with different initial sending rates $x_f(0)$ along with the optimal rates. The results show that the system stabilizes around the optimal rate, independent of the initial condition. The small fluctuations around the optimal value are caused by the imprecise channel measurements and the communication delay between the flow sources and the intermediate nodes where prices are generated.

Fig. 6.9 shows that the value of $\kappa$ will affect the stability and the convergence rate of the algorithm. In particular, if $\kappa$ is too large (*e.g.*, $\kappa = 0.005$), the flow rates always fluctuate. The value of $\kappa$ needs to be small enough ($< 0.002$) to ensure the stabilization of the system. On the other hand, the algorithm will converge slower with smaller $\kappa$ value.

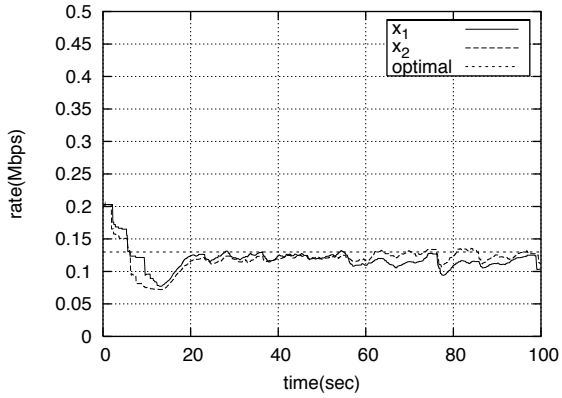
We now proceed to show the fairness of our price-based rate allocation. Recall that when the utility function is set to $U_f(x_f) = \log(x_f)$, the price-based rate allocation is able to achieve proportional fairness [17]. Here we present some intuitive properties of proportional fairness. (1) If flows $f_1$ and $f_2$ share the same path (uses the same amount of resources), then $x_1 = x_2$; (2) if flow $f_1$ uses more bottleneck resources than $f_2$, then $x_1 < x_2$. Specifically, if the prices of flow $f_1$ and $f_2$ are $\kappa_1$ and $\kappa_2$, then $\frac{x_1}{x_2} = \frac{\kappa_2}{\kappa_1}$. To quantitatively study the property of fairness, we define the fairness index in the single resource element case (no spatial reuse) as $I = \frac{(\sum_{f \in F}(x_f/H)^2}{|F| \times \sum_{f \in F}(x_f/H)^2}$, where $H$ is the number of hops the flow passes in this channel. Note that $I \in [0, 1]$. Larger value of $I$ indicates better fairness. Fig. 6.10 plots the fairness index in the simulation scenario as shown in Fig. 6.7. We compare our algorithm with TCP, which is shown to be unfair for end-to-end flows in wireless backhaul mesh network [15]. The results show that our algorithm outperforms TCP in terms of fairness, independent of the values of $\beta$ and $\tilde{cw}$.

## 6.6 Related Work

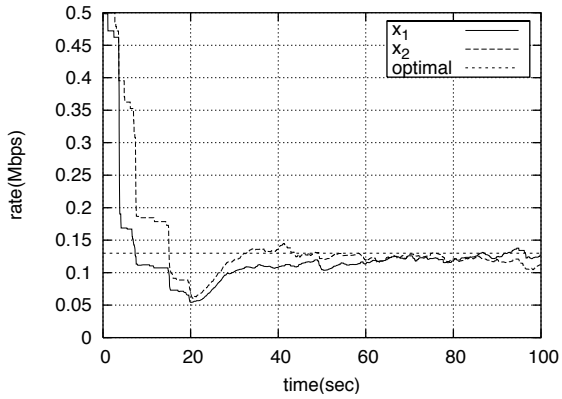
We compare and highlight the contributions of this work in light of previous related work.




(a)

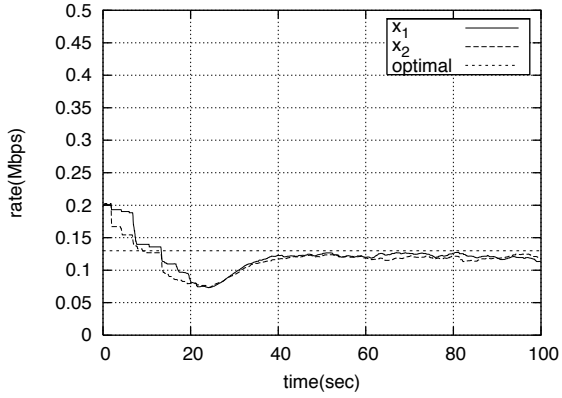



(b)

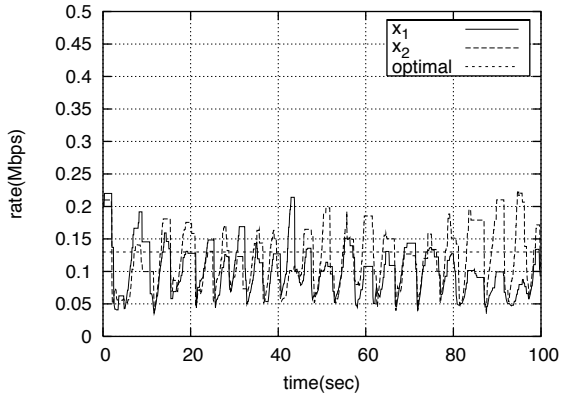



(c)

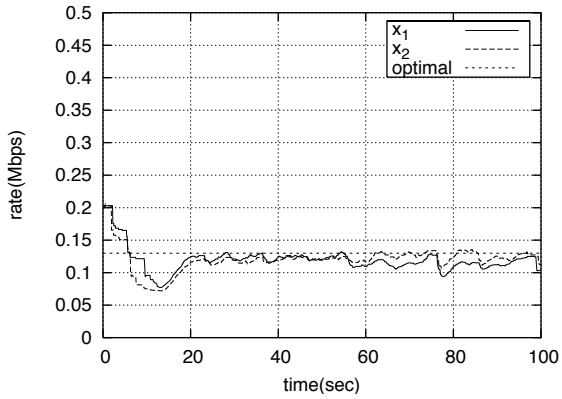
**Fig. 6.8.** Instantaneous throughput with different initial sending rates: (a) $x(0)$ = 20 Kbps, (b) $x(0)$ = 200 Kbps, and (c) $x(0)$ = 500 Kbps.




(a)

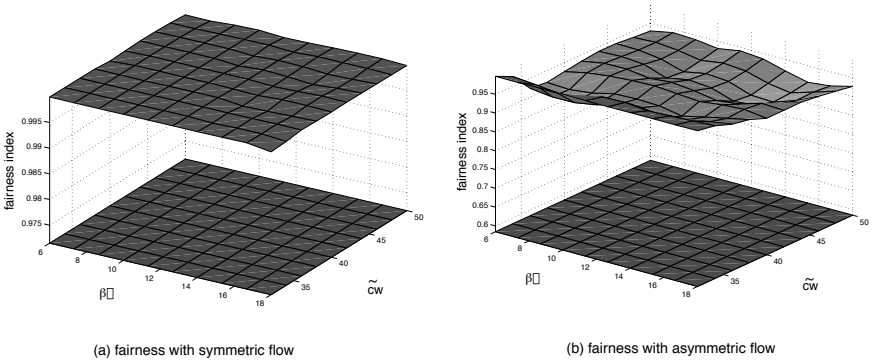



(b)




(c)

**Fig. 6.9.** Instantaneous throughput with different values of $\gamma$: (a) $\gamma = 0.005$, (b) $\gamma = 0.002$, and (c) $\gamma = 0.001$.




(a) fairness with symmetric flow

(b) fairness with asymmetric flow

**Fig. 6.10.** Fairness.

Existing research on WMN has focused on how to better utilize the wireless channel resource and enhance its performance. Proposed solutions include equipping mesh nodes with multiple radios and distributing the wireless backbone traffic over different channels, routing the traffic through different paths [18, 19], or a joint solution of these two [20,21]. These existing approaches usually fall into two ends of the spectrum. On one end of the spectrum are the heuristic algorithms (*e.g.*, [18,21]). Although many of such approaches are adaptive to the dynamic environments of wireless networks, they lack the theoretical foundation to analyze how well the network performs globally (*e.g.,* whether the network resource is fully utilized, whether the flows share the network in a fair fashion). On the other end of the spectrum, there are theoretical studies that formulate these network planning decisions into optimization problems (*e.g.*, [22,23]). Yet these results usually make ideal assumptions and present centralized algorithms. None of them has realistically considered the highly dynamic and distributed nature of wireless mesh network environments. Further, these existing solutions only apply to routing and channel allocation, while our work addresses the resource allocation and rate limiting problem for wireless mesh network.

The problem of optimal and fair resource allocation has been extensively studied in the context of wireline networks, where pricing has been shown to be an effective approach (*e.g.,* [9, 24, 25]). Our approach is similar to [9, 25], which solves the resource allocation problem using a penalty-based approach. Nevertheless, the fundamental differences in contention models between multihop wireless and wireline networks deserve a fresh treatment to this topic. One of the highlights of this work is to propose a generalized theoretical framework of resource allocation that fits both wireline and multihop wireless networks. Within this framework we show that the resource model of wireline network is just a special case, in comparison with multihop wireless networks.

The problem of fair and effective resource allocation in multihop wireless networks has also been previously studied, using MAC-layer fair scheduling which targets on *single-hop* MAC layer flows [26]- [28]. In comparison, this work studies

*end-to-end* multihop flows in such networks. It can be shown that fair resource allocation among single-hop flows may not be optimal for multi-hop flows, due to the unawareness of bottlenecks and lack of coordination among upstream and downstream hops. Moreover, global optimal resource allocation among multi-hop flows can not be completely reached only by MAC-layer scheduling, which is only based on local information. In this context, the only remedial solution is to use prices as signals to coordinate global resource allocation.

This work is also related to the work of [29] and [11], in that both works explore the fundamental performance limit of a multihop wireless network in presence of interference. Yet our work is different from these works in the following aspects. First, we seek to maximize the aggregated utility, which can be a nonlinear function of flow rates. Such an objective can achieve maximum throughput under different fairness models by specifying appropriate utility functions. Second, the solution space of their approach is defined by routing [11] or joint scheduling and routing [29], while this work defines its solution space by rate control. Third, the works of [11,29] aim at deriving the limit of optimal throughput by centralized algorithms. This work focuses on how to achieve such a limit, by presenting a decentralized algorithm and a distributed implementation.

The work of [30] provides an intuitive solution to improve TCP fairness via neighborhood RED. Our work can be regarded as its theoretical interpretation. Our early work in [16] also presents a resource allocation algorithm for multihop wireless networks. Compared to [16], this chapter presents a more precise and general resource model from the view of independent set polytope. Moreover, in terms of resource allocation algorithm, [16] uses a dual approach which provides an exact solution to the original resource allocation problem directly. In contrast, the resource allocation framework presented in this paper uses a primal approach that solves the relaxation of the original problem. We argue that this approach may be more suitable for real deployment in multihop wireless mesh networks, as the prices can be generated directly from channel conditions.

A collection of papers have studied the use of price in the context of wireless networks (*e.g.*, [31, 32]). In these papers, pricing has been used as a mechanism for optimal distributed power control. In addition, Liao *et al.* [33] use prices to provide incentives for service allocation in wireless LANs. The work of [34, 35] also uses prices as incentives to encourage packet relays in multihop wireless networks. Our work is different from these works in that we apply pricing to regulate resource usage rather than providing incentives.

## Conclusion

This chapter targets on the resource allocation problem in wireless mesh network. What we desire is a generalized theoretical framework, which can effectively capture the common nature of these problems, *i.e.*, they can all be categorized as constrained non-linear optimization problem. Applying this generalized framework to the setting of multihop wireless mesh backbone network, we find out that a resource element is

not a wireless link, but a facet of the polytope determined by independent sets in the conflict graph of the wireless network. Through this finding, we are able to outline the solution space of the resource allocation problem in wireless mesh network, and derive the corresponding decentralized algorithm. The same framework also provides theoretical evidence to help judge the feasibility of existing solutions. We reveal that the fundamental problem of TCP unfairness in multihop wireless networks lies in its incorrect congestion signal (*i.e.*, price). Our work also offers theoretical validation to recent-proposed solutions, such as neighborhood RED [30], IFA [15].

## References


1. "Mesh networks inc.," http://www.meshnetworks.com.
2. "Radiant networks," http://www.radiantnetworks.com.
3. "Seattle wireless," http://www.seattlewireless.net.
4. "MIT roofnet," http://www.pdos.lcs.mit.edu/roofnet/.
5. "Chaska wireless solutions," http://www.chaska.net/.
6. R. Karrer, A. Sabharwal, and E. Knightly, "Enabling large-scale wireless broadband: The case for taps," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 27–32, 2004.
7. J. Bicket, D. Aguayo, S. Biswas, and R. Morris, "Architecture and evaluation of an unplanned 802.11b mesh network," in *Proc. of ACM MobiCom*, 2005.
8. F. P. Kelly, "Charging and rate control for elastic traffic," *European Trans. on Telecommunications*, vol. 8, pp. 33–37, 1997.
9. F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.
10. P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, pp. 388–404, 2000.
11. K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact on interference on multi-hop wireless network performance," in *Proc. of ACM MobiCom*, 2003.
12. M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, *Springer*, 2000.
13. J. G. Augustson and J. Minker, "An analysis of some graph theoretical cluster techniques," *Journal of ACM*, vol. 17, no. 4, pp. 571–586, 1970.
14. K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact on interference on multi-hop wireless network performance," in *Proc. of ACM Mobicom*, 2003.
15. V. Gambiroza, B. Sadeghi, and E. W. Knightly, "End-to-end performance and fairness in multihop wireless backhaul networks," in *Proc. of ACM MobiCom*, 2004.
16. Y. Xue, B. Li, and K. Nahrstedt, "Optimal resource allocation in wireless ad hoc networks: A price-based approach," *IEEE Transactions on Mobile Computing*, vol. 5, no. 4, pp. 347–364, April 2006.
17. S. Kunniyur and R. Srikant, "End-to-end congestion control: Utility functions, random losses and ECN marks," in *Proc. of IEEE INFOCOM*, 2000.
18. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. of ACM Mobicom*, 2004.
19. Y. Yuan, H. Yang, S. H. Y. Wong, S. Lu, and W. Arbaugh, "Romer: Resilient opportunistic mesh routing for wireless mesh networks," in *Proc. of IEEE WiMesh*, 2005.

20. A. Raniwala, K. Gopalan, and T. Chiueh, "Centralized channel assignment and routing algorithms for multi-channel wireless mesh networks," *Mobile Computing and Communications Review*, vol. 8, no. 2, pp. 50–65, 2004.
21. A. Raniwala and T. Chiueh, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. of IEEE INFOCOM*, 2005.
22. M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. of ACM MobiCom*, 2005.
23. M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. of IEEE WiMesh*, 2005.
24. S. H. Low and D. E. Lapsley, "Optimization flow control, I: Basic algorithm and convergence," *IEEE/ACM Trans. on Networking*, vol. 7, no. 6, pp. 861–874, 1999.
25. S. Kunniyur and R. Srikant, "End-to-end congestion control: Utility functions, random losses and ECN marks," in *Proc. of IEEE INFOCOM*, 2000.
26. H. Luo, S. Lu, and V. Bharghavan, "A new model for packet scheduling in multihop wireless networks," in *Proc. of ACM Mobicom*, 2000.
27. L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless networks," in *Proc. of IEEE INFOCOM*, 2002.
28. Y. Liu and E. Knightly, "Opportunistic fair scheduling over multiple wireless channels," in *Proc. of IEEE INFOCOM*, 2003.
29. M. Kodialam and T. Nandagopal, "Characterizing the achievable rates in multihop wireless networks," in *Proc. of ACM Mobicom*, 2003.
30. K. Xu, M. Gerla, L. Qi and Y. Shu, "Enhancing TCP fairness in ad hoc wireless networks using neighborhood RED," in *Proc. of ACM Mobicom*, 2003.
31. T. M. Heikkinen, "On congestion pricing in a wireless network," *Wireless Networks*, vol. 8, no. 4, pp. 347–354, 2002.
32. D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," in *Proc. of IEEE INFOCOM*, 2002.
33. R. Liao, R. Wouhaybi, and A. Campbell, "Incentive engineering in wireless LAN based access networks," in *Proc. of IEEE ICNP*, 2002.
34. Y. Qiu and P. Marbach, "Bandwidth allocation in ad-hoc networks: A price-based approach," in *Proc. of IEEE INFOCOM*, 2003.
35. L. Buttyan and J. P. Hubaux, "Stimulating cooperation in self-organizing mobile ad hoc networks," *ACM/Kluwer Mobile Networks and Applications*, vol. 8, no. 5, October 2003.

# 7

# Resource Allocation and Cost in Hybrid Solar/Wind Powered WLAN Mesh Nodes

A. A. Sayegh, T. D. Todd, and M. N. Smadi

McMaster University, Canada
todd@mcmaster.ca

## 7.1 Introduction

WLAN mesh networks are currently being deployed for outdoor wireless coverage in many metro-area Wi-Fi hotzones. One of the costs of these mesh deployments is that of providing nodes with continuous electrical power. In many of these cases, continuous node powering from the AC power mains is expensive or practically impossible. This is increasingly true as network coverage moves into more expansive outdoor areas.

An alternative to a fixed power connection is to operate some of the WLAN nodes using a sustainable energy source such as solar or wind power. The SolarMESH network is an operational testbed deployment which uses this approach [1]. In either solar or wind powered options, node resource allocation involves assigning solar panel or wind turbine size, and battery capacity to each mesh node. This assignment must use "geographic provisioning" to account for the solar insolation or wind power capability of the node location. Resource assignment is done using a target load profile for the node, which specifies the power consumption workload for which it is being configured. Since the cost of the battery and the solar panel or wind turbine can be a significant fraction of the total node cost, it is important that power consumption on the node is minimized as much as possible.

In this chapter we present geographic provisioning results for solar and wind powered WLAN mesh nodes. A cost model is introduced which is used to optimize the hybrid provisioning of the nodes. The results suggest that in certain geographic locations a hybrid wind/solar powered WLAN mesh node is the optimum cost configuration. Cases will be included using existing IEEE 802.11 standard assumptions and will also consider the case where modifications are made to the standard so that mesh AP power saving is possible. Several North American locations have been chosen for these results, i.e., Toronto and Yellowknife, Canada; Seattle, Wa. and Phoenix, Az. These locations have been chosen to illustrate a variety of differing meteorological situations.

## 7.2 Background

Solar panel and battery sizing methods were studied in [2] along with discussions of how various photovoltaic (PV) components result in different system configurations. Three methods for sizing PV systems were compared in [3]. The first assigns the battery capacity such that it can support a fixed load for a preset number of days. Then the solar panel is sized so that a full battery re-charge can be done within a specified time period. The second scheme is based on computer simulations using historical solar insolation data which is used to track the evolution of the battery state-of-charge. In the third approach [4] a Markovian model was used for battery state of charge modeling. To use this method however, the mean and variance of the daily solar insolation must be known. This model was refined in [5] and accounts for the effects of daily correlation in the solar irradiation. Another performance evaluation of PV systems based on Markovian modeling was presented in [6]. In [3] it was shown that the simulation method yields the most accurate results.

In [7] and [8] the sky clearance index was used for simulating solar irradiance, and in [9] variable loading was taken into account in PV system sizing. A fuzzy decision process was used in [10] for evaluating subjective factors in PV system sizing. Energy management in a space station was also considered assuming limited energy constraints [11].

Much of the literature assumes an idealized battery model which can sometimes lead to significant inaccuracies. In practical systems the battery capacity is a strong function of ambient temperature and must be taken into consideration. In [12, 13] and [14] models for non-ideal battery behavior were studied.

In [15] a study was presented of the resource allocation problem in solar powered WLAN mesh nodes. This work also showed the potential of mesh node power saving on the resource assignment. It was shown that there can be a significant reduction in node cost when AP power saving is used. In addition, several outage control strategies were presented that can be used to prevent mesh node outage.

A number of previous works investigated the use of hybrid powered systems [16]-[19]. In [16], long-term hourly weather data for 30 years was used in order to calculate the optimum size of a PV array for a stand-alone hybrid wind/PV system. This was done by generating probability density functions of the wind speed and solar insolation for each hour in any given month. A least squares method was then used in order to find the best fit of the PV array and wind turbine for a given load. In [17] a similar approach was used, except that a more refined method was introduced for calculating the probability density function. This was then used to compute the number of required storage batteries and PV panels. Readers who are interested in the theoretical background of wind and solar power characteristics should refer to [20], which provides a step-by-step analysis of both energy sources.

In [18] a simple numerical algorithm was used to compute the optimum generation capacity and storage needed for three scenarios in a remote area in Montana, with a typical residential load. These options are stand-alone wind, PV, and hybrid wind/PV. The paper then performs an economic analysis in order to measure the cost effectiveness of the three scenarios. Finally, in [19] the output power of wind tur-

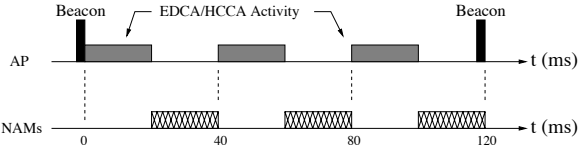



**Fig. 7.1.** Best-effort mesh AP power saving with movable boundary.

bines in cold weather was discussed. The conclusions were based on experiments performed in Tiverton, Canada, and showed that wind turbines are practical in cold regions.

Many of the above studies deal with the tradeoffs between node resource requirements and the power consumption of the node. The ability for WLAN mesh infrastructure to conserve power is a highly desirable capability and can lead to significantly reduced node cost. This issue is discussed in detail in the next section.

### 7.2.1 Power Saving in WLAN Mesh Access Point

WLAN mesh APs which operate using a sustainable energy source can benefit greatly from reduced power consumption [15]. IEEE 802.11 however, does not provide a mechanism for placing APs into a power saving mode. The standard does define power saving for client devices, by allowing the nodes to switch between the Awake and Doze states. While in the Awake state, the client device is fully powered and is capable of transmitting, receiving, and of sensing channel activity. Conversely, during the Doze state, the client device operates in power save (PS) mode where it is not able to perform any of these functions. When the station is in this state the AP to which it is associated must hold any incoming packets until the client returns to the Awake state. In an IEEE 802.11 AP, beacon packets are transmitted periodically, which include a traffic indicating map (TIM), indicating if packets are currently being buffered on behalf of PS mode stations. The retrieval of packets from the AP is then done one at a time using PS-Poll frames transmitted by the station. A notification of pending broadcast and multicast traffic is indicated in a delivery traffic indication map (DTIM). The delivery of the actual downlink broadcast and multicast traffic is done immediately after the beacons are sent.

IEEE 802.11e has a power saving mechanism that is similar to that of legacy IEEE 802.11 and includes both contention and polling based options. For example, the Hybrid Coordinator (HC) can define periodic service intervals which allow the synchronous delivery of traffic using Automatic Power Save Delivery (APSD).

To realize the cost savings associated with power saving mesh APs, two variations of IEEE 802.11 have recently been proposed [21, 22]. In [21], the AP dynamically modifies its sleeping schedule to adaptively support current loading conditions while saving as much power as possible. Reference [22] proposed a power saving WLAN mesh architecture based on the IEEE 802.11e [23].

In [22] a power saving AP includes a network allocation map (NAM) in its beacon broadcasts, which specifies periods of time within the superframe when it is

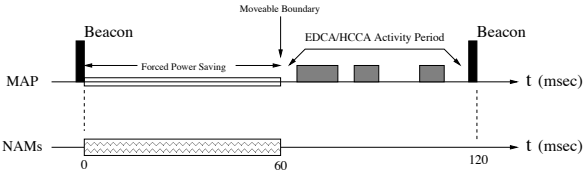



**Fig. 7.2.** Forced mesh AP power saving (FPS) with 50% offered capacity.

unavailable. During these periods the AP is assumed to be inactive and conserving power. Fig. 7.1 shows an example of this type of activity for a single inter-beacon period. The channel activity is shown in the upper timeline, and the NAMs are shown on the lower timeline. In this example three HCCA (or EDCA) periods (each lasting 20 ms) have been scheduled and the AP advertises the NAMs as shown so that power saving can occur when the channel is not needed. The example shown might be used when the AP is supporting a combination of VoIP connections (based on a 40 ms packetization interval) and best-effort data traffic. In this case the AP is satisfying the quality of service requirements of the mobile stations, and using the remaining time for power saving. An algorithm for dynamically updating the NAMs was proposed in [22].

In some cases it is desirable for the AP to force a maximum level of activity, regardless of mobile station traffic requirements. This may happen when the AP workload exceeds that for which the node was originally provisioned, leading to the possibility of node outage. Rather than an outage, the AP may artificially reduce its offered capacity, and this is referred to as Forced Power Saving (FPS). Fig. 7.2 shows an example of an AP that is using FPS. In this example the AP advertises a NAM restricting its activity to a maximum of 50% of the inter-beacon interval. This capability can be used to develop outage control algorithms as proposed in [15].

WLAN mesh point relay links which do not require communication with standard compliant end devices may use proprietary power saving mechanisms. This approach is currently being considered by the IEEE 802.11s working group, based on modified versions of conventional IEEE 802.11 protocols [24].

## 7.3 Hybrid Sustainable Energy WLAN Mesh Nodes

In this section we discuss the basic configuration of a hybrid powered WLAN mesh AP (MAP) or mesh point (MP) node. A simplified block diagram is shown in Fig. 7.3. The system includes a wind turbine and/or solar panel, and a battery. When both a solar panel and wind turbine are used, we refer to this as a *hybrid* configuration. The solar panel and/or wind turbine are connected to the battery through a charge controller which disconnects the battery to protect it from under- and over-charging.

We can define an energy flow model for this configuration, where $\mathrm{e}_{panel}(k)$ is the energy produced in the solar panel over the time increment $[(k-1)\Delta, k\Delta]$, and $\Delta$ is the time-step length considered. In hybrid PV systems designed using publicly

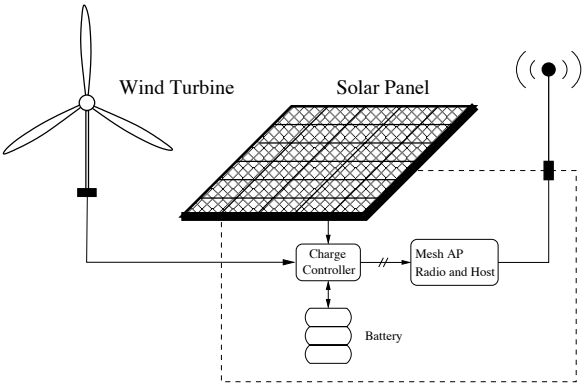



**Fig. 7.3.** Solar/wind powered WLAN mesh node components.

available meteorological data, data collection and modeling is done in discrete time, using 1 hour $\Delta$ increments. The solar panel size is given by $S_{panel}$, and is normally rated in watts at peak solar insolation. $\mathrm{e}_{turbine}(k)$ is the energy produced by the wind turbine and it is a function of wind speed W, i.e.,

$$\mathrm{e}_{turbine}(k) = \frac{1}{2}\xi\rho\pi R^2 W^3 \Delta \tag{7.1}$$

where $\xi$ is the efficiency of the turbine, which cannot exceed the theoretical limit of 59.26%. This restriction was discovered by Betz in 1919. In practice, the achievable wind turbine output power is much lower than this limit due to losses in the alternator and due to the non-ideal aerodynamic properties of the turbine blades. Typical values for commercial wind turbines are in the 30% range. In (7.1), $\rho$ represents the air density which is roughly 1.23 kg per cubic meter at sea level. Wind turbines have two additional key parameters, namely, the *cut-in speed* which is the minimum wind speed at which the turbine starts generating power, and the *cut-out speed* where the turbine must be turned away from the wind to protect the blades.

We also define $\mathcal{B}(k)$ to be the residual battery energy stored at time $k\Delta$, and $\mathbf{B}_{max}$ is defined to be the total battery capacity. If we assume that $\mathcal{L}(k)$ is the load energy demand over the time duration $[(k-1)\Delta, k\Delta]$, then we can write [6]

$$\mathcal{B}(k) = \min\{\max[\mathcal{B}(k-1) + \mathrm{e}_{panel}(k) + \mathrm{e}_{turbine}(k) - \mathcal{L}(k), \mathbf{B}_{outage}], \mathbf{B}_{max}\}. \tag{7.2}$$

In the above equation, $\mathbf{B}_{outage}$ is the maximum allowed depth of discharge, based on safety and battery life considerations [2]. When $\mathcal{B}(k) < \mathbf{B}_{outage}$, the charge controller will disconnect the MAP/MP load and the node will experience a radio outage. It is also important to take into account the temperature because any reduction leads to a reduced charge storage capability in the battery. This effect has been taken into account in the results that will be presented later.

In most photovoltaic applications, fixed solar panels are pointed directly south and sloped slightly greater than the geographic latitude so that solar absorption is

highest during winter months. Meteorological data, however, is only available for horizontal and fully-tracking (direct normal) components and cannot be used directly for a fixed planar solar panel. For this reason a conversion model is used to compute the energy incident on the solar panel using horizontal and fully-tracking irradiance records. In [15] different conversion methodologies are shown for the solar data. In the remainder of this chapter, we assume that the solar data has been properly converted to match the installed panel. The value of the solar insolation is rated in watts, and it is normalized to the panel's peak power. For example, when the insolation is 0.1 and it is incident on a 60 W solar panel, the overall power generated is 6 W.

## 7.4 Energy Source Examples

In this section we give some example configurations of renewable energy resource assignment using the above procedures for two diverse climatic locations, i.e., Toronto, Canada, and Phoenix, Az. For a given location it is important to quantify the joint statistical distribution of the wind and the solar energy. For example, if wind speed increases greatly during nighttime hours and is almost zero during the day when there is bright sunlight, this would suggest that a hybrid wind/solar system might be cost effective. In this case a node could operate during the daytime hours from the solar panel and at night from the wind turbine thus minimizing the size of the battery needed for either case alone. The same argument applies to long term correlations between the energy sources.

Our simulator implements the energy balance equation shown in (7.2) for a given geographic location. It also accepts as input the publicly available meteorological information for a given geographic location. These records are maintained in the National Solar Radiation Data Base (NSRDB), National Renewable Energy Laboratory (NREL), U.S. Department of Energy [25] for the USA while in Canada data may be obtained from the National Climate Data and Information Archive, The Meteorological Service of Canada (MSC) [26]. The data is usually presented on an hourly basis, i.e., $\Delta = 1$ hour. The simulator incorporates the algorithm mentioned in [15] to convert the available solar data based on the different solar angles. The simulator provides the outage events and the battery charge at every hour during the simulation. This approach is similar to that adopted by several commercial simulators such as PVSYST and PV-Design Pro.

### 7.4.1 Example Short-Term Statistics

In this section, we investigate the short-term statistics of the different energy sources for two example locations, Toronto, Canada, and Phoenix, Az. In the results, we assume that we use a small, commercially available wind turbine of fixed size, the Muartec Rutland 503 [27] which generates approximately 24 watts at a wind speed of 10 m/s. For the city of Toronto, we assume a 60 W solar panel which corresponds to the wind energy available at the maximum wind speed.

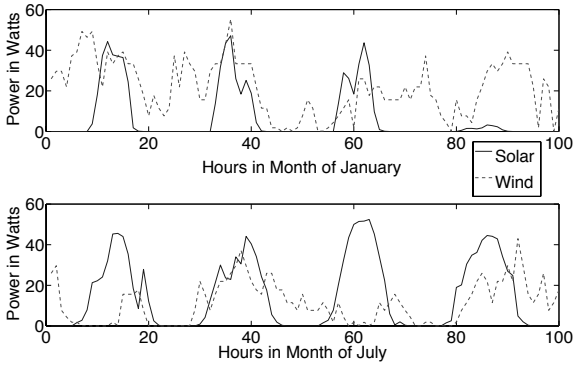



**Fig. 7.4.** Comparison of solar power and wind power for Toronto in January and July.

Fig. 7.4 shows a time distribution example of solar power and wind power for the first 100 hours in January and July, 1990, for the city of Toronto. It can be seen that in January, the wind power is more dominant than solar power. We also observe a strong positive correlation between the increase in solar insolation and the increase in wind speed, yet the power available from the wind is much greater. For example, if we examine the first 24 hours, we see that the wind power is always present and it peaks at almost 50 W while the solar insolation is not always available and peaks at 42 W.

The situation is reversed in July where we see that the solar insolation outperforms the wind power. For example, the solar insolation peaks during the day at slightly less than 45 W while at the same time the wind dies down to values less than 18 W during the day. During the night, we again see a very strong correlation, i.e, both sources produce negligible power.

The positive solar/wind power correlation that is observed for Toronto would suggest that a hybrid solar/wind powered node may not be cost effective. When the sources are strongly correlated, it may be best to use the one that is the most cost effective. However, this conclusion needs to be verified by incorporating a cost model which will influence the optimal mix of both energy sources. This will be treated in more detail in Section 7.6.

If we consider the city of Phoenix, the maximum wind power is 35 W. In this case we compare to a solar panel with peak power equal to 35 W. Fig. 7.5 shows the power distribution over time for the first 100 hours of January and July, 1990. We can see that the solar power clearly dominates the wind power for January and July, indicating that it may always be more cost-effective. We can see that the collected solar power is almost always at its peak of 35 W (and zero at night), while the wind power rarely exceeds 10 W. In addition, we notice that even though the increase in wind speed is correlated to the decrease in solar insolation, the values are so low that it may not significantly reduce the required battery size. These traces suggest that in a location such as Phoenix, wind power alone or a hybrid wind/solar solution is probably not feasible.

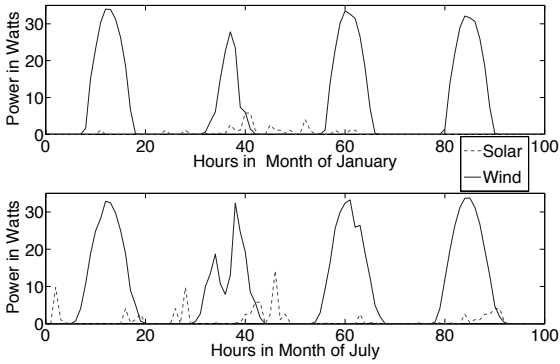



**Fig. 7.5.** Comparison of solar power and wind power for Phoenix in January and July.

### 7.4.2 Example Long-Term Statistics

We now consider some examples of the long-term behavior of the wind and the solar energy sources. In the first set of results we present examples that compare the relative value of solar versus wind power, when the total average power from each source is the same. For comparison purposes we assume the source can generate a long term average power output of 2 W, which is roughly the minimum power consumption of a single radio WLAN mesh AP whose radio is always active [1].

For the city of Toronto (1990) the average wind speed is 7 m/s, and in order to supply an average of 2 W, we would need a 7 cm wind turbine assuming a turbine efficiency of 30%. On the other hand, the average solar insolation is 0.1746, so the solar panel size would be approximately 11.5 W. The comparison is shown in Fig. 7.6. We can see that for a city like Toronto with a temperate climate the solar powered node slightly outperforms the wind-powered node. For example, at a battery size of 10 Ah the wind outage is 0.28 while it is 0.2 for the solar-powered node. For Phoenix, the average wind speed is 2.6945 and therefore, the wind turbine radius would be 35 cm. The solar insolation is 0.2651, which yields a solar panel size of 7.5443 W. These results are also shown in Fig. 7.6. We can see that the solar panel in Phoenix greatly outperforms the wind turbine. For example, at a battery size of 50 Ah, the solar panel outage is almost zero while it is still 0.2 for the wind turbine case.

The behavior seen in Fig. 7.6 is somewhat counter-intuitive. If the long-term average output power from solar and wind power is the same, and if both processes were stationary over short time periods (on the order of days), then one might expect that wind power would achieve better outage performance than solar. This is because solar insolation is always absent at night time, whereas the same is not true for wind power. On this basis one could argue that solar insolation is more "bursty" than wind power (for the same long term average), and thus requires a higher battery capacity to achieve the same outage probability.

The above behavior is explained by considering the seasonal correlation between solar and wind power. In Fig. 7.7 we show the state of charge of an initially full battery when it is powered only by a wind turbine and then only by solar power. Again

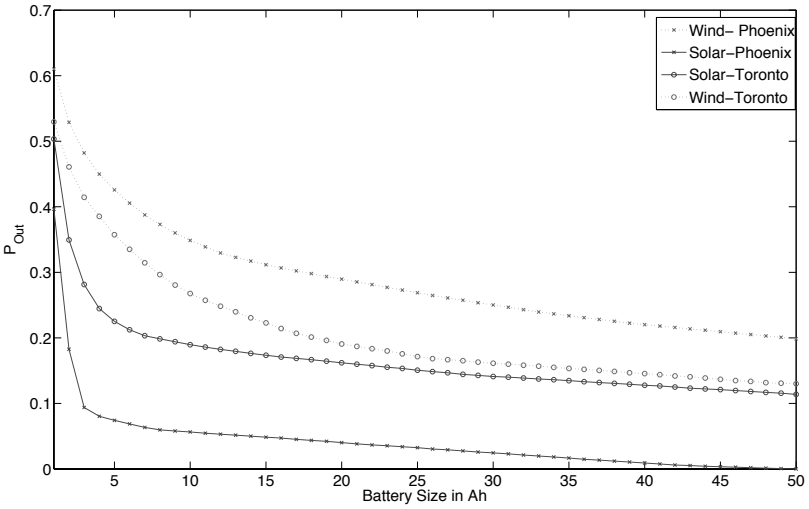



**Fig. 7.6.** Comparison of sources for Toronto and Phoenix for a 2 W load with the same average power.

both sources produce an average of 2 W. Examining the curves carefully, we can see that the wind and solar energy have a strong negative seasonal correlation. For example, in the summer the battery is always full when powered by the sun while it is almost empty when powered by the wind. The situation is reversed to a lesser degree in late fall and winter when the wind energy increases while the solar energy is greatly reduced. These results strongly suggest that a node placed in Toronto would benefit from a hybrid design with more emphasis on solar energy, however, the ratios of the contributions of the sources will be dictated by cost considerations. The short-term statistics would have suggested that wind power is more perpetual on an hourly basis especially at night when the solar panel generates zero energy. However, after examining these results it is clear that the wind and solar energy for Toronto have the potential for augmenting each other on a seasonal basis. These observations cannot be generalized, i.e., if we examine Fig. 7.8 we can see a similar comparison of sources for Phoenix. It is clear that the node almost never runs into an outage when powered by solar energy. On the other hand, we can see that the wind energy fluctuates seasonally, leading to outage. In this case it seems that without considering the relative costs involved, solar energy is the clear winner as the node's energy supply.

In Fig. 7.9 for Seattle, the results are similar to Toronto, though both the wind and solar energy are slightly less. Finally, for Yellowknife, as shown in Fig. 7.10, we can see that the solar energy is very weak and almost nonexistent in the fall and winter. The wind power during this period is better, which suggests that the wind source would be useful in supplementing the node during the darker months. Again, the results suggest a hybrid approach.

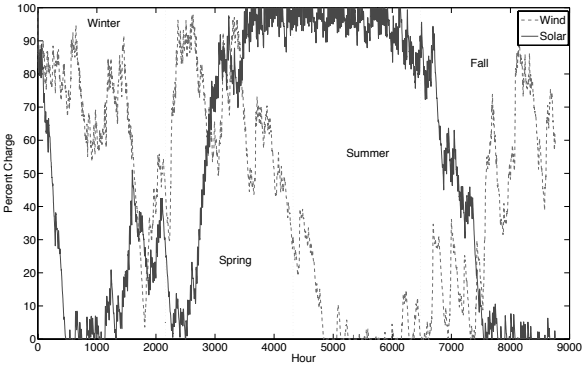



**Fig. 7.7.** Percentage charge versus time for different sources for Toronto with a 50 Ah battery.

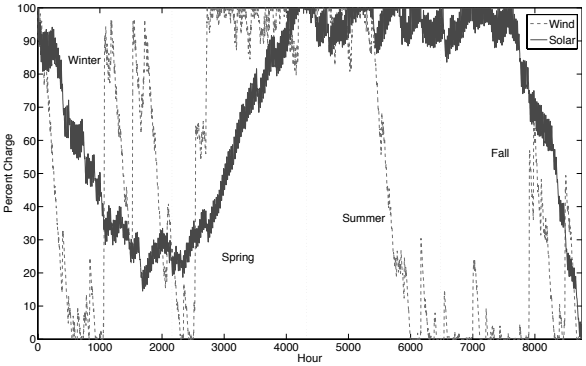



**Fig. 7.8.** Percentage Charge versus Time for Different Sources for Phoenix with a 50 Ah Battery.

## 7.5 Node Sustainability

In this section we consider the sustainability of a given hybrid system for the different geographical locations. We consider two possible scenarios. In the first case, we assume a small commercial wind turbine (Muartec Rutland 503) that is used to supplement the system. The other scenario to be considered is when the turbine size may be arbitrarily chosen. It is important to point out that the performance of any practical wind turbine is less than the theoretical optimum. This effect is shown in Fig. 7.11. However, it should be noted that at the most common wind speeds the values of the power for both cases are almost identical. In all of the following simulations we assume that the air density is constant and equal to 1.23 $kg/m^3$ and that the cut-in and cut-out speeds are $cut_{in} = 3.75$ m/s, $cut_{out} = 20$ m/s, respectively. We also assume that the load to be powered is a constant 2 W.

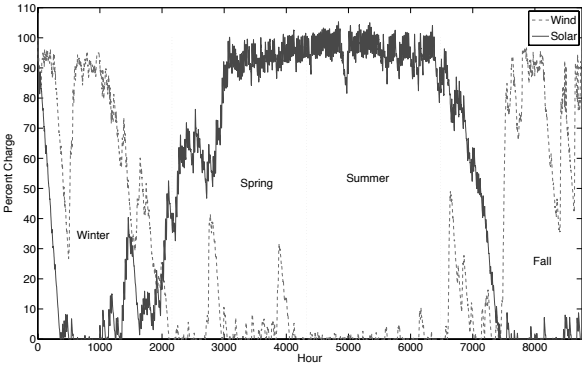



**Fig. 7.9.** Percentage charge versus time for different sources for Seattle with a 50 Ah battery.

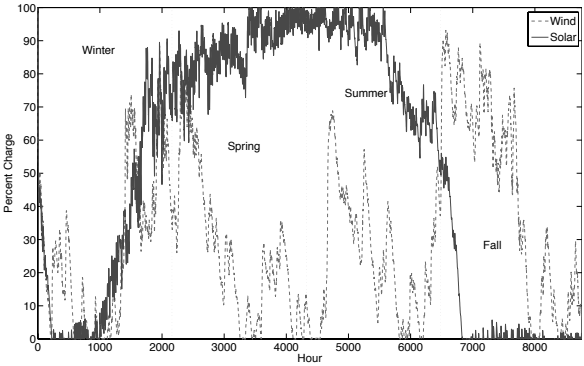



**Fig. 7.10.** Percentage charge versus time for different sources for Yellowknife with a 50 Ah battery.

### 7.5.1 Fixed Wind Turbine Source

In this section we assume a specific wind turbine configuration (the Muartec Rutland 503) with no solar panel, and simulate the system to find the battery sizes needed to eliminate outage. This wind turbine is one of the few that are commercially available at about the size needed for WLAN mesh AP applications. Fig. 7.12 shows the outage probability versus the battery size for the 4 different geographical locations. We can see that a battery size of 12.5 Ah will completely eliminate outages for Toronto, and for the other cities a much larger battery is needed and even then it may not eliminate outages completely. By comparison, the same turbine installed at Phoenix will never eliminate outages even for a very large battery due to the scarcity of wind power as previously discussed. We also see that the results for Yellowknife and Seattle are between those for Toronto and Phoenix. We notice that Yellowknife greatly outperforms Seattle which indicates that the wind energy is more abundant in that region even though the storage capabilities of the battery are greatly impaired due to cold temperatures effects.

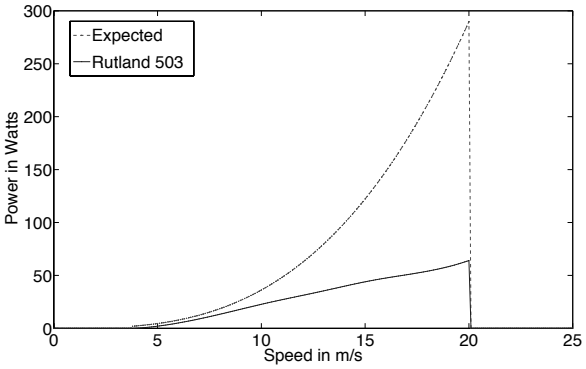



**Fig. 7.11.** Theoretical vs. real performance of turbine.

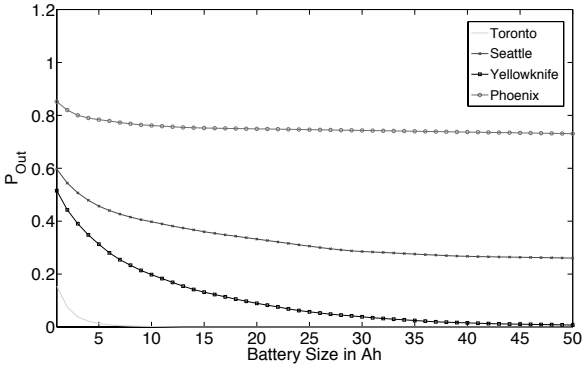



**Fig. 7.12.** $P_{Out}$ vs. battery size for the Muartec Rutland 503 wind turbine.

It can be seen that the option of simply installing a commercially available wind turbine will not eliminate outages in some cases even for very large battery sizes. In the following section, we examine the case when we can choose the size of the turbine freely without the restriction of current commercial availability.

### 7.5.2 Variable Wind Turbine Size

In this section we assume that we can select the wind turbine size for a given location. Fig. 7.13 shows the contour plots of the outage probability for different turbine radii and battery sizes for Toronto. For example, if an outage probability of 0.0001 is required, the battery size will be 15 Ah with a turbine radius of 15 cm. From this we can conclude that wind power in Toronto can successfully eliminate outage. However, as we will see in the next section, it is important to take into account the cost of the turbine. In addition, examining the seasonal correlation with solar power will dictate the optimal mix of both sources.

On the other hand, Fig. 7.14 shows the corresponding contour plots for Phoenix. Clearly the resources needed at Phoenix are much higher than Toronto. By following

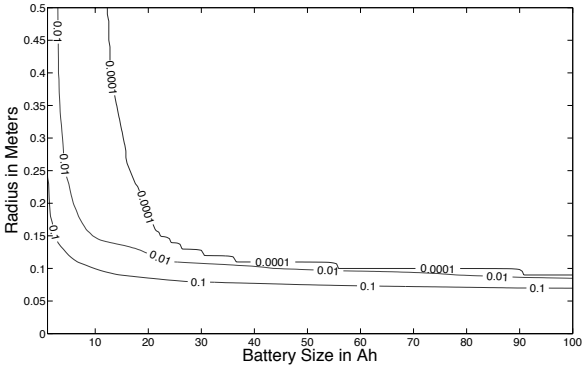



**Fig. 7.13.** Contour plot of $P_{Out}$ for different wind turbine and battery sizes for Toronto.

the same example, the battery size would be 45 Ah which is a threefold increase. The wind turbine radius is 70 cm which is an increase of 4.5 times over that of Toronto. We can conclude that installing a wind turbine at Phoenix is likely to be prohibitively expensive and will never successfully eliminate outages.

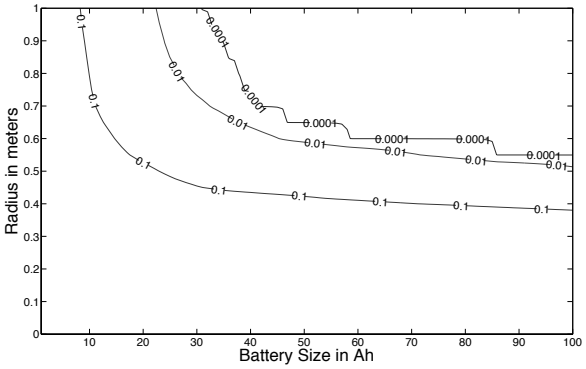



**Fig. 7.14.** Contour plot of $P_{Out}$ for different wind turbine and battery sizes for Phoenix.

Fig. 7.15 shows the results for Seattle, and we can see that the battery size is 25 Ah and the turbine radius is 30 cm. Finally, Fig. 7.16 shows the results for Yellowknife, where the required battery size is 70 Ah and the turbine radius is 35 cm. We can conclude that in most locations, wind energy will be sufficient to meet an acceptable outage probability criterion, however this will be at the expense of the costs associated with the size of the wind turbine and battery needed.

In the following section, we examine the effect of integrating the cost model with the energy model for each city in order to obtain the optimal mix of energy sources and to find out when it would be best to deploy a hybrid node versus a node powered by a single energy source.

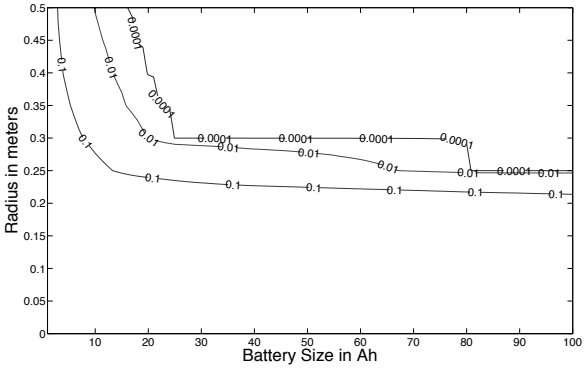



**Fig. 7.15.** Contour plot of $P_{Out}$ for different wind turbine and battery sizes for Seattle.

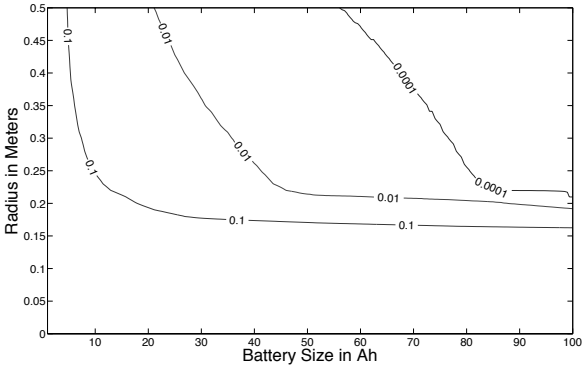



**Fig. 7.16.** Contour plot of $P_{Out}$ for different wind turbine and battery sizes for Yellowknife.

## 7.6 Hybrid Node Cost Optimization

From the previous discussion, we can conclude that in most cases the addition of a wind turbine will greatly improve the performance of a WLAN mesh node. However, the addition of the turbine must justify its added cost to the system. In this section we provide a cost optimization for a hybrid solar/wind node. The optimization must incorporate a realistic model in order to determine the cost-optimal resource allocation. In the following we develop a cost model which aims to be as realistic as possible. We assume current (2007 year) retail values. We assume that the costs include the installation costs but do not account for differences in the ongoing maintenance of the solar or wind powered components. Also, the fixed cost of the node electronics is not included. For the wind turbine, the Canadian Wind Energy Association (CanWea) [28] states that the cost of a small wind turbine is in the range of 5000-6400 CAD/KW. KW ratings are usually at 10 m/s for 300 W to 1000 W. Commercially available turbines are not very common at the ratings we are considering due to the lack of demand, and because of the use of wind technology for much higher powered

applications. For this reason, we mainly rely on extrapolating the costs from larger wind turbines.

Based on publicly available values, we can see that the relationship between the power output and the turbine area can be written as 1 KW = 0.28 $\times Area$ CAD, and hence we can write that $Cost \approx 1400 \times \pi \times R^2$ CAD, where R is the turbine radius. Therefore, the cost of the wind turbine is roughly proportional to the square of the radius. For experimentation, we introduce an economy-of-scale discount factor for smaller radii, $\eta$. This factor will be used to experiment with the optimal target costs for the wind turbine. Using commercial data sheets for the solar panel, we find that the cost is roughly $\alpha \times P$ CAD, where $P$ is the peak panel power in watts, $\alpha$ is typically about 6.7. Similarly, the cost of a lead acid battery is about $\beta \times B$ CAD where $B$ is the battery size in Ah and $\beta$ is typically about 3.4.

Since the average power is also a quadratic function of the wind turbine radius, the cost per watt is constant for the turbine. In Toronto, for example, it is about 10.45 CAD per watt and is constant for any radius. This is calculated for the average power generated by a turbine for the year 1990. On the other hand the cost per watt for the solar panel is about 41 CAD. This is due to the fact that the average power from a panel is low since it only produces power during the day and no power during the night. For Phoenix it is 25.27 CAD/W for the solar panel and 245.83 CAD/W for the wind turbine. For Seattle it is 46.3 for the panel and 63.51 for the wind turbine. For Yellowknife it is 40.98 for the panel and 59.95 for the wind turbine.

We wish to minimize the total cost of the node while making sure that outage does not exceed a target outage probability, $P_{Outdesign}$. We also assume that the battery, solar panel and turbine have an upper and lower bound on size and that we will only optimize over discrete values of $B$, $P$ and $R$. Our optimization problem is given as

$$\min_{B,P,R} \mathit{Cost} \tag{7.3}$$

such that

$$Cost = Cost_P + Cost_B + Cost_R \tag{7.4}$$

$$Cost_P = \alpha P \tag{7.5}$$

$$Cost_B = \beta B \tag{7.6}$$

$$Cost_R = 4400 \times \eta R^2 \tag{7.7}$$

$$P_{Out} = f(B, P, R) \tag{7.8}$$

$$P_{Out} \leq P_{Outdesign} \tag{7.9}$$

$$B_{Min} \leq B \leq B_{Max} \tag{7.10}$$

$$P_{Min} \leq P \leq P_{Max} \tag{7.11}$$

$$R_{Min} \leq R \leq R_{Max}. \tag{7.12}$$

In (7.3), we seek to minimize the total cost of the node. The cost function is comprised of the sum of three components as seen in 7.4; namely, the costs of the battery, the solar panel and the wind turbine as seen in (7.5), (7.6), and (7.7). These cost components follow the cost model that we previously discussed. In this optimization,

**Table 7.1.** Parameter definitions used in the optimization.

| Parameter | Definition |
|---|---|
| $\alpha$ | Cost per unit solar panel power. |
| $\beta$ | Cost per unit battery storage. |
| $\eta$ | Economy of scale factor for the wind turbine. |
| $P_{Out}$ | Outage probability. |
| B | Battery size in Ah. |
| P | Solar panel size in watts. |
| R | Wind turbine radius in meters. |

$P_{Out}$ is defined as the outage probability of the node. Unfortunately, $P_{Out}$ is a complex non-linear function of $B$, $P$ and $R$ which we define as $f(B, P, R)$ as seen in 7.8. This non-linearity makes the optimization problem very difficult. In (7.9) there is a constraint on $P_{Out}$ since it must satisfy a design target, referred to as $P_{Outdesign}$. Finally, the constraints in (7.10), (7.11), and (7.12) specify that there are upper and lower bounds on $B$, $P$, and $R$. Since discrete values of $B$, $P$, and $R$ are required, the above problem can be classified as an integer programming problem. The parameters defined above are listed in Table 7.1.

Due to the non-linearity of $P_{Out}$, we adopt an iterative approach to solving this problem. First, we perform a discrete event simulation of different battery/panel/turbine configurations using the tool developed in [15], where we assume the increment in battery and solar panel sizes is 1 Ah and 1 W respectively, and for the turbine radius we assume increments of 1 cm. The simulation will provide the associated outage probability for a given geographic location and node configuration. Once the simulation is completed, we are then able to use the results of a discrete optimizer that we wrote using Matlab that finds the optimal values of $B$, $P$, and $R$ for a given geographic location.

We assume the outage probabilities, 0, 0.0001, 0.001, 0.01 and 0.1, for $P_{Outdesign}$. We also assume that $\alpha = 6.7, \beta = 3.4$, and $\eta = 1$. Finally, we have assumed that $B_{max} = 50$, $P_{max} = 50$, and $R_{max} = 0.5$. The results are shown in Tables 7.2, 7.3, 7.4, and 7.5 and apply for a 2W load for the year 1990.

As shown in Table 7.2, for Phoenix the most cost effective configurations do not include the wind turbine and hence $R = 0$ for all cases. This is not surprising based on our previous discussion of the solar and wind energy available at that location. If the required load is increased to 4 W, we can see in Table 7.6 that still the wind turbine size is zero while the total cost has increased in a linear fashion. If we consider Phoenix with a discount factor of $\eta = 0.25$ as shown in Table 7.7, we still see that there is no value in using a wind powered source.

As seen in Table 7.3 for Toronto, a wind turbine of radius 0.1 m is always necessary in order to achieve optimal cost. This is due to the high cost of the battery and due to its reduced storage capacity due to cold temperatures. By comparison if we examine Table 7.8, we can see the effect of eliminating the wind turbine on the total node cost. For example, for the case of $P_{Out} = 0.1$ the cost is reduced from

146.53 CAD to 74.2 CAD. If we examine Toronto for higher loads of 4 W and 10 W as shown in Tables 7.9 and 7.10 respectively, we can see that for the 4 W case, the turbine size does not increase much. However, at the 10 W load level the turbine size increases to 23 cm for the zero outage case. As expected the cost of the node increases with the increase in loading, following an almost linear trend. If we consider a discount factor of 0.5 for a load of 2 W, Table 7.11 shows that the size of the wind turbine increases by 50% to 15 cm.

For Seattle, as shown in Table 7.4, we see that a turbine of radius 0.11 is required for minimum cost and decreases to 0.1. In all cases, more resources need to be allocated than for Toronto. Finally, if we examine the Yellowknife location (Table 7.5), we see that a hybrid approach works well. We notice that the resources required for Yellowknife are much higher than the other locations we have considered.

If we consider the zero outage case, we can see that the total cost of the system for Yellowknife is 338.57 CAD, in Toronto it is 131.6 CAD, in Seattle it is 242.23 CAD, and in Phoenix it is 107.2 CAD. The cost of deploying a node in Yellowknife is over three times the cost that provides the same level of service in Phoenix. By comparison, the cost in Yellowknife is almost two and a half times the cost in Toronto and it is slightly less than one and a half times the cost in Seattle. We also observe that relaxing the constraints on the outage probability leads to significant cost savings, for example, in Toronto, relaxing the outage target from 0 to 0.1 leads to a reduction in cost from 131.6 CAD to 71.69 CAD, which is almost a factor of 2. On the other hand, in Phoenix the reduction is from 107.2 CAD to 63.53 CAD and in Yellowknife it is reduced from 338.75 CAD to 175.03 CAD. Therefore, for Toronto, Seattle, and Yellowknife a hybrid solar/wind powered node is the most optimum from a cost viewpoint. This is due to the fact that even though solar power is scarce, wind turbines are very expensive and hence a trade-off is necessary in order to minimize the cost.

**Table 7.2.** Minimum cost node configuration for Phoenix (2 W load).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 8 | 12 | 0 | 0 | 0 | 107.2 |
| 8 | 12 | 0 | 0 | 0.0001 | 107.2 |
| 8 | 11 | 0.02 | 0.0009 | 0.001 | 102.29 |
| 6 | 10 | 0.02 | 0.0099 | 0.01 | 88.83 |
| 3 | 8 | 0 | 0.075 | 0.1 | 63.53 |

### 7.6.1 Mesh AP Power Saving

In this section, we examine the effect of using mesh AP power saving on the cost of the WLAN mesh node. Tables 7.12, 7.13, 7.14, and 7.15 show the results of simulations with a load power of 0.5 W. Examining Table 7.12 we can see that the required

**Table 7.3.** Minimum cost node configuration for Toronto (2 W load).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 14 | 60 | 0.1 | 0 | 0 | 131.6 |
| 140 | 60 | 0.1 | 0 | 0.0001 | 131.6 |
| 130 | 60 | 0.1 | 0.0009 | 0.001 | 128.2 |
| 140 | 30 | 0.0900 | 0.0099 | 0.01 | 103.24 |
| 30 | 50 | 0.0800 | 0.0951 | 0.1 | 71.69 |

**Table 7.4.** Minimum cost node configuration for Seattle (2 W load).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 33 | 10 | 0.12 | 0 | 0 | 242.23 |
| 33 | 10 | 0.12 | 0 | 0.0001 | 242.23 |
| 32 | 10 | 0.12 | 0.0005 | 0.001 | 238.83 |
| 21 | 13 | 0.11 | 0.0098 | 0.01 | 211.31 |
| 8 | 11 | 0.10 | 0.0961 | 0.1 | 144.53 |

**Table 7.5.** Minimum cost node configuration for Yellowknife (2 W load).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 39 | 14 | 0.16 | 0 | 0 | 338.57 |
| 39 | 14 | 0.16 | 0 | 0.0001 | 338.57 |
| 37 | 14 | 0.16 | 0.0008 | 0.001 | 331.77 |
| 30 | 13 | 0.16 | 0.0095 | 0.01 | 301.31 |
| 10 | 10 | 0.13 | 0.0994 | 0.1 | 175.03 |

**Table 7.6.** Minimum cost node configuration for Phoenix for a 4 W load.

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 17 | 23 | 0 | 0 | 0 | 211.13 |
| 17 | 23 | 0 | 0 | 0.0001 | 211.13 |
| 17 | 22 | 0 | 0.00079 | 0.001 | 204.67 |
| 14 | 19 | 0 | 0.0095 | 0.01 | 174.27 |
| 6 | 15 | 0 | 0.0927 | 0.1 | 120.4 |

resources are greatly reduced from 131.6 CAD to almost 34 CAD for the zero outage case. By comparison, the reduction is from 107.2 to 26.8 CAD for Phoenix, from 242.3 to 62.09 CAD for Seattle, and finally, from 338.57 to 85.43 CAD for Yellowknife. We can see that for all cities the total cost is reduced almost linearly by a factor of 4. This is due to the fact that the total cost is proportional to the average power load of the node. These examples illustrate the cost and size reductions possible when power saving is implemented.

**Table 7.7.** Minimum cost node configuration for Phoenix for $\eta = 0.25$.

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 9 | 11 | 0.04 | 0 | 0 | 105.69 |
| 9 | 11 | 0.04 | 0 | 0.0001 | 105.69 |
| 8 | 11 | 0.02 | 0.0009 | 0.001 | 100.97 |
| 6 | 10 | 0.02 | 0.0099 | 0.01 | 87.51 |
| 3 | 8 | 0 | 0.075 | 0.1 | 63.53 |

**Table 7.8.** Minimum cost node configuration for Toronto with no wind turbine.

| Battery Size, B | Solar Panel Size, P | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|
| 29 | 26.99 | 0 | 0 | 278.59 |
| 36.12 | 23 | 0.0001 | 0.0001 | 276.15 |
| 29 | 25.96 | 0.001 | 0.001 | 271.69 |
| 24 | 24.93 | 0.01 | 0.01 | 247.82 |
| 10 | 16.88 | 0.1 | 0.1 | 146.53 |

**Table 7.9.** Minimum cost node configuration for Toronto (4 W load).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 38 | 13 | 0.1 | 0 | 0 | 259.87 |
| 38 | 13 | 0.1 | 0 | 0.0001 | 259.87 |
| 35 | 12 | 0.11 | 0.0009 | 0.001 | 252.24 |
| 26 | 8 | 0.12 | 0.0098 | 0.01 | 205.09 |
| 8 | 9 | 0.11 | 0.0937 | 0.1 | 140.44 |

**Table 7.10.** Minimum cost node configuration for Toronto (10 W load).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 50 | 43 | 0.23 | 0 | 0 | 689.43 |
| 50 | 43 | 0.23 | 0 | 0.0001 | 689.43 |
| 49 | 49 | 0.19 | 0.0008 | 0.001 | 652.11 |
| 48 | 27 | 0.20 | 0.0098 | 0.01 | 519.20 |
| 25 | 15 | 0.19 | 0.0997 | 0.1 | 343.84 |

## Conclusion

In this chapter, we have presented geographic provisioning results for solar and wind powered WLAN mesh nodes. A cost model has been introduced which is used to optimize the provisioning of such networks. The model suggests that in certain geographic locations a hybrid wind/solar powered WLAN mesh node is the optimum cost configuration. The presented results have compared various design alternatives including infrastructure power saving and non-power saving options.

**Table 7.11.** Minimum cost node configuration for Toronto (2 W load) $\eta = 0.5$.

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 14 | 0 | 0.15 | 0 | 0 | 97.10 |
| 14 | 0 | 0.15 | 0 | 0.0001 | 97.10 |
| 13 | 0 | 0.15 | 0.0007 | 0.001 | 93.70 |
| 8 | 0 | 0.15 | 0.0075 | 0.01 | 76.70 |
| 6 | 0 | 0.11 | 0.0925 | 0.1 | 47.02 |

**Table 7.12.** Minimum cost node configuration for Toronto for a 0.5W load (power-saving).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 4 | 2 | 0.04 | 0 | 0 | 33.97 |
| 4 | 2 | 0.04 | 0 | 0.0001 | 33.97 |
| 4 | 2 | 0.04 | 0 | 0.001 | 33.97 |
| 4 | 1 | 0.04 | 0.0066 | 0.01 | 27.31 |
| 2 | 1 | 0.04 | 0.0437 | 0.1 | 20.51 |

**Table 7.13.** Minimum cost node configuration for Phoenix for a 0.5W load (power-saving).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 2 | 3 | 0 | 0 | 0 | 26.80 |
| 2 | 3 | 0 | 0 | 0.0001 | 26.80 |
| 2 | 3 | 0 | 0 | 0.001 | 26.80 |
| 2 | 3 | 0 | 0 | 0.01 | 26.80 |
| 1 | 2 | 0 | 0.06 | 0.1 | 16.73 |

**Table 7.14.** Minimum cost node configuration for Seattle for a 0.5W load (power-saving).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 8 | 2 | 0.07 | 0 | 0 | 62.09 |
| 8 | 2 | 0.07 | 0 | 0.0001 | 62.09 |
| 8 | 2 | 0.07 | 0 | 0.001 | 62.09 |
| 6 | 3 | 0.06 | 0.01 | 0.01 | 56.24 |
| 2 | 3 | 0.05 | 0.08 | 0.1 | 37.80 |

As an example, for the city of Toronto our results suggest that the cost increases greatly with the load power. The cost for the node at zero outage increased from 33.97 to 131.6 to 259.87 to 689.43 CAD when the load increased from 0.5 to 2 to 4 to 10 W. This example has showed the importance of power-saving on the node cost. Our results have also showed that for Toronto a hybrid solution is much more cost effective than a solar powered approach, where the cost was reduced from 278.59 to 131.6 CAD. Our results have also shown that not all geographic locations will be able to make use of a hybrid mix of energy sources. For example, locations such as

**Table 7.15.** Minimum cost node configuration for Yellowknife for a 0.5W load (power-saving).

| Battery Size, B | Solar Panel Size, P | Wind Turbine Radius, R | $P_{Out}$ | $P_{Outdesign}$ | Cost |
|---|---|---|---|---|---|
| 9 | 4 | 0.08 | 0 | 0 | 85.43 |
| 9 | 4 | 0.08 | 0 | 0.0001 | 85.43 |
| 9 | 4 | 0.08 | 0 | 0.001 | 85.43 |
| 7 | 4 | 0.08 | 0.0071918 | 0.01 | 78.63 |
| 3 | 2 | 0.07 | 0.092580 | 0.1 | 45.09 |

Phoenix, Az cannot make use of wind power to reduce cost due to the abundance of solar insolation. We have also shown that the temporal distribution of the power sources is of utmost importance when the node is being sized. Our results suggest that the short-term statistics are not sufficient in order to assess the optimal ratio of solar to wind power used in the system. The long-term and yearly statistics are much more important.

Finally, our results and examples have shown that mesh AP power saving is highly beneficial since it reduces the allocated resources and hence the node cost. This is based on the observation that total cost is almost linearly proportional to the node load power consumption which may be greatly reduced by using power saving.

## References

1. The SolarMESH Network. http://owl.mcmaster.ca/solarmesh/. McMaster University. Hamilton, Ontario, Canada., 2004.
2. L. Narvarte and E. Lorenzo, "On the usefulness of stand-alone PV sizing methods," *Progress in Photovoltaics: Research and Applications, Prog. Photovolt: Res. Appl.*, 8:391–409, 2000.
3. H. A. M. Maghraby, M. H. Shwehdi, and G. K. Al-Bassam, "Probabilistic assessment of photovoltaic (PV) generation system," *IEEE Transactions on Power Systems*, 17(1):205–208, 2002.
4. L. L. Bucciarelli, "The effect of day-to-day correlation in solar radiation on the probability of loss of power in a stand-alone photovoltaic energy system," *Solar Energy*, 36(1):11–14, 1986.
5. L. L. Bucciarelli, "Estimating loss-of power probabilities of stand-alone photovoltaic solar energy systems," *Solar Energy*, 32(2):205–209, 1984.
6. F. M. Safie, "Probabilistic modeling of solar power systems," in *Proc. of Annual Reliability and Maintainability Symposium, 1989.*, pages 425–430, 1989.
7. I. Abouzahr and R. Ramakumar, "Loss of power supply probability of stand-alone photovoltaic systems: A closed form solution approach," *IEEE Transactions on Energy Conversion*, 6(1):1–11, 1991.
8. U. Grasselli, "Probabilistic design of high quality power supply photovoltaic systems," in *Industrial and Commercial Power Systems Technical Conference, 1993. Conference Record, Papers Presented at the 1993 Annual Meeting*, 1993.

9. S. Saengthong and S. Premrudeepreechacham, "A simple method in sizing related to the reliability supply of stand-alone photovoltaic systems," in *Proc. of the Twenty-Eighth IEEE Photovoltaic Specialists Conference, 2000.*, pp. 1630–1633, 2000.
10. D. Macomber, "Optimizing residential photovoltaic system size using approximate reasoning," in *Proc. of First International Symposium on Uncertainty Modeling and Analysis*, pp. 558–563, 1990.
11. M. Bouzguenda and S. Rahman, "Energy management onboard the space station-a rule-based approach," *IEEE Transactions on Aerospace and Electronic Systems*, 27(2):302–310, 1991.
12. P. E. Baikie, M. I. Gillibrand, and K. Peters, "The effect of temperature and current density on the capacity of lead-acid battery plates," *Electrochimica Acta*, 17:839–844, 1972.
13. Z. M. Salameh, M. A. Casacca, and W. A. Lynch, "A mathematical model for lead-acid batteries," *IEEE Transactions on Energy Conversion*, 7(1):93–98, 1992.
14. A. Pesaran and V. Johnson, "Battery thermal models for hybrid vehicle simulations," *Journal of Power Sources*, 110:377–382, 2002.
15. A. Farbod, "Design and resource allocation for solar-powered ESS mesh networks," Master's thesis, McMaster University, 1280 Main St. West, Hamilton, Ontario, Canada L8S 4K1, August 2005.
16. B. S. Borowy and Z. M. Salameh, "Optimum photovoltaic array size for a hybrid wind/PV system," *IEEE Transactions on Energy Conversion*, 9(3):482–488, 1994.
17. A. D. Bagul, Z. M. Salameh, and B. Borowy, "Sizing of a stand-alone hybrid wind-photovoltaic system using a three-event probability density approximation," *Solar Energy*, 56(4):323–335(13), April 1996.
18. W. Kellogg, M. Nehrir, G. Venkataramanan, and V. Gerez, "Generation unit sizing and cost analysis for stand-alone wind, photovoltaic, and hybrid wind/PV systems," *IEEE Transactions on Energy Conversion*, 13(1):70–75, March 1998.
19. C. Leclerc and C. Masson, "Abnormally high power output of wind turbine in cold weather: A preliminary study," *International Journal of Rotating Machinery*, 9(1):23–33, 2003. doi:10.1155/S1023621X03000034.
20. D. Heinemann, "Energy Meteorology: Lecture Notes," *Postgraduate Programme Renewable Eenergy Carl von Ossietzky University*, 2002.
21. F. Zhang, T. D. Todd, D. Zhao, and V. Kezys, "Power saving access points for IEEE 802.11 wireless network infrastructure," in *Proc. of IEEE Wireless Communications and Networking Conference 2004 (WCNC'04)*, March 2004.
22. Y. Li, T. D. Todd, and D. Zhao, "Access point power saving in solar/battery powered IEEE 802.11 ESS mesh networks," in *Proc. of the Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE'2005)*, August 2005.
23. IEEE Standards Department. *Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements*. IEEE Press, 2005.
24. IEEE Standards Department. 802.11s ESS Mesh Networking working group, 2004.
25. National Solar Radiation Data Base. http://rredc.nrel.gov/solar/. National Renewable Energy Laboratory (NREL), U.S. Department of Energy, 2004.
26. National Climate Data and Information Archive. http://www.climate.weatheroffice.ec.gc.ca/. The Meteorological Service of Canada, Canada, 2004.
27. Marlec Engineering Co Ltd. "Rutland 503 Wind Charger Data Sheet". *http://www.marlec.co.uk/products/prods/rut503.htm*.

28. Canadian Wind Energy Association CanWEA. “Cost Comparison for Small Wind Turbine Sizes”. *http://www.smallwindenergy.ca/en/Overview/Costs/CostComparison.html*.

# 8

# Scheduling, Routing, and Related Cross-Layer Management through Link Activation Procedures in Wireless Mesh Networks

L. Badia[1], A. Erta[1], L. Lenzini[2], and M. Zorzi[3]

[1] IMT Lucca Institute for Advanced Studies, Italy
{l.badia, a.erta}@imtlucca.it
[2] University of Pisa, Italy
l.lenzini@iet.unipi.it
[3] University of Padova, Italy
zorzi@dei.unipd.it

## 8.1 Introduction

In a Wireless Mesh Network (WMN) [1] end users are provided with wireless broadband connectivity by means of a pre-defined system hierarchy. To describe this organization, several notations can be used. In the following, we adopt the terminology of [2]. The end terminals, also referred to as Mesh Clients (MCs), are connected to special nodes, denoted as Mesh Routers (MRs). These nodes do not generate traffic, since they are simply meant to relay the packets of their MCs. Additionally, some MRs, called Mesh Access Points (MAPs), can be provided with a wired connection, and can therefore act as gateways toward the Internet. The MAPs are also wirelessly interconnected to all the other MRs in a multi-hop fashion, without necessarily following pre-defined paths. Instead, an MC can interact only with the MR it is connected to. MRs form what is usually named as the *backbone* of the WMN, which can physically cover a large region in a wireless manner. This structure offers a good cost/benefit balance, since it almost entirely avoids cable set up. For this reason, it is deemed to be applicable in rural areas, where the deployment of wireline networks may be too expensive. WMNs can also be envisaged for dense residential or business areas, and in general, anyplace where the installation of cables is difficult because of physical obstacles.

There are several possibilities to specify the Medium Access Control (MAC) used by a WMN. These are often related to existing standards, especially IEEE 802.11 [3] and IEEE 802.16 [4], parts of which are dedicated to WMNs. Actually, the first hop from any MC to its related MR is often assumed to employ a radio access interface different from the one used in the backbone, and entirely *orthogonal* (i.e., perfectly non-interfering) to it, e.g., since it uses another frequency, and possibly another technology. Moreover, the first hop may adopt management strategies typical

of cellular networks [5], and is therefore conceptually simpler. For this reason, we will not investigate this part of the WMN in greater detail. Conversely, realizing the interconnections among MRs poses many theoretical challenges, most of which are common to all kinds of multi-hop networks, such as ad hoc and sensor networks. However, when revising them for WMNs, some important properties come into play. Usually, MCs can be portable devices, whereas MRs and MAPs are not mobile. Therefore, the backbone management does not suffer from most mobility issues, neither at the transport layer (i.e., paths do not need to be updated), nor at the physical layer (channel variability is relatively moderate). Moreover, communications in a WMN are usually to or from the Internet, thus all routes have either the source or the destination in a MAP. Finally, as MRs can be easily placed near to a power outlet, energy saving is not an issue. These properties considerably distinguish the backbone of WMNs from an ad hoc network (for what concerns pre-defined hierarchy and absence of mobility) or a sensor network (lack of terminal battery limitations).

The issues which arise in the backbone management relate to different layers of the protocol stack. On the one hand, the creation of low-interference and high-rate paths to the MAPs is key to achieve good rates at each MR. This may also involve the exploitation of multiple channels as, for example, MRs can own several Network Interface Cards (NICs), which can simultaneously operate on different frequencies. On the other hand, the link layer needs to schedule packets over multiple links in order to achieve good transmission parallelism and possibly forward more data towards the MAPs at the same time.

The main problems which will be investigated by our analysis are:

- routing algorithms, i.e., network level procedures to discover efficient paths which connect the ordinary MRs (and therefore their MCs) to the MAPs. Note that routing strategies designed for ad hoc networks usually admit also peer-to-peer communications, which are not common for WMNs. Moreover, the goal in WMNs is more often to obtain high system throughput rather than maximizing battery lifetime.
- link scheduling, which involves medium access level procedures to activate communication links. Its goal is to ensure network connectivity while at the same time satisfying physical constraints related to technology, interference and network management.
- cross-layer management, operating at an intermediate level with both network and link layer procedures, jointly addressing these problems.

The aforementioned issues involve other related topics, which are also worth discussing. In certain cases very broad subjects are involved, which will be discussed here only for what concerns their impact on the definition of routing and scheduling strategies. There are also other aspects of these matters which fall out of the scope of the present article, and therefore, will not be discussed here in detail. However, the reader will be addressed to external references to find further material on them. Some of the related problems which will be framed into our analysis are:

- channel assignment and node placement: in our analysis, these are considered to be aspects of network deployment, which means they have already been per-

formed at the time routing or scheduling strategies are sought. However, it will be briefly outlined how it is possible to incorporate them into the same cross-layer framework used for routing and scheduling with a modular approach, thus with no need for significant modifications of the reasonings presented in the rest of the article.

- models of wireless interference: for this point, two important considerations must be made. First of all, we propose a detailed review and classification of the possible approaches to characterize interference. We try to resolve terminology ambiguity due to the use of different names for the same model or of the same name for distinct models in the literature. Moreover, we discuss the choice of the model itself, which is driven by two contrasting aspects. On the one hand, the interference model should be as accurate as possible. In this sense, the use of heavily simplified interference models may end up in poor algorithm performance when applied to realistic cases. On the other hand, a certain degree of approximation is unavoidable as related to the properties of the Medium Access Control (MAC) protocol. In fact, in a layered network management, algorithms operating on top of the link layer necessarily abstract some aspects of the physical layer, such as interference. For these reasons, we will first concentrate our analysis on general results which hold true regardless of the interference model, such as theoretical performance bounds. Then, we will discuss how these findings translate to practical cases, at which point different interference models need to be taken into account.

The rest of this chapter is organized as follows. In Section 8.2 we give a brief overview of the problem studied and we clarify terminology and notations employed in the rest of the chapter. In Section 8.3 we present a review of the works which discussed related topics in a way applicable to WMNs. In Section 8.4 we mathematically formalize the problem, in particular identifying the constraints determined by capabilities of the terminals and wireless interference. This latter aspect, in particular, is discussed in depth in Section 8.5, proposing a classification of interference models, and also touching MAC protocol issues. In Section 8.6 we give both theoretical and practical evaluations of the performance of WMNs. Even though the problem is NP-complete and exact approaches are hard, we present some original analytical results which determine both upper and lower performance bounds, and we give quantitative insights by applying them to sample WMN topologies. Finally, we present the conclusions.

## 8.2 Preliminaries

We represent the backbone of a WMN as a graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. The *nodes* in set $\mathcal{N}$ are the MRs, which are in turn connected by the *edges* belonging to set $\mathcal{E} \subseteq \mathcal{N}^2$, thus representing the communication links of the backbone. This approach is commonly used for multi-hop wireless networks [6,7], even though the graph is often considered bi-directional, i.e., with undirected edges. This is a limiting assumption,

as will be discussed in Subsection 8.5.1. Similar to [8, 9], we will assume instead that the edges, as actual wireless communication links, are uni-directional. Thus, the communication link where a sender node $i \in \mathcal{N}$ transmits to a receiver $j \in \mathcal{N}$ is represented by an element $e \in \mathcal{E}$ equal to the ordered pair $(i, j)$. The inclusion of this link in $\mathcal{E}$ actually happens only if node $j$ can receive a transmission from $i$ in the absence of any other interference source.

In the following, we will denote with $\mathcal{R}_i$ and $\mathcal{S}_i$ the set of nodes which are possible receivers from and senders to node $i$, respectively. In other words, $\mathcal{R}_i$ and $\mathcal{S}_i$ contain the one-hop output and input neighbors of $i$. Formally:

$$\mathcal{R}_i = \{j \in \mathcal{N} : (i, j) \in \mathcal{E}\} \tag{8.1}$$

$$\mathcal{S}_i = \{j \in \mathcal{N} : (j, i) \in \mathcal{E}\}. \tag{8.2}$$

We will also refer to other properties of the communication link represented by edge $(i, j) \in \mathcal{E}$. To quantify the *capacity* of the link we make use of variables $r_{ij}$, called *link rates* and collected into a matrix $\mathbf{R} = (r_{ij})$. Rate $r_{ij}$ can be regarded as the number of bits which can be transmitted over the link represented by edge $(i, j)$ in a given time unit. When required by physical specifications, we will also consider a parameter $g_{ij}$ corresponding to the wireless link gain[1] over $(i, j)$. A matrix $\mathbf{G} = (g_{ij})$ can be introduced collecting the $g$ variables for all edges.

In our investigations, we consider an underlying Space and Time Division Multiple Access (STDMA) scheme [10]. For wireless multi-hop networks it is in fact crucial to exploit space and time parallelism in order to obtain an efficient transmission scheme.

Our mathematical representation of scheduling and routing over the WMN backbone is similar to the ones reported in [6,8,11]. A link represented by edge $(i, j) \in \mathcal{E}$ is said to be *active* if node $i$ transmits to node $j$. Thus, for any edge $e \in \mathcal{E}$ of the graph, we define a binary variable $x_e(t)$, which varies over a discrete (slotted) time and indicates activation of the corresponding link at time $t$, i.e., $x_e(t) = 1$ if the link is active and $x_e(t) = 0$ otherwise. By varying $t$, the activation variables $x_e(t)$ determine a time-division scheduling for the WMN backbone according to what we refer to in the following as *link activation pattern*. Similar to the analysis presented in [11–13], we remark that the derivation of the scheduling through a link activation pattern implicitly determines the routing as well. This is visible, for example, in Fig. 8.1, where a packet needs to be sent from A to D. Assume that nodes B and C do not have packets to send themselves and can act as relays. A route is created from node A to node D by subsequently allocating links $e$ (from A to B), $f$ (from B to C), and $g$ (from C to D). Note that the entire route is actually realized by operating over three time slots.

To be efficient, such an STDMA link activation scheme needs to be aware of the network topology. This is a strong requirement in many types of wireless multi-hop networks, where nodes are mobile, but as the backbone usually consists of

[1] The wireless link gain is the ratio between received and transmitted power. It is well known that wireless channels are strongly time-varying. However, for simplicity, we will consider slowly varying scenarios where the $g_{ij}$ parameters can be approximated as constants.

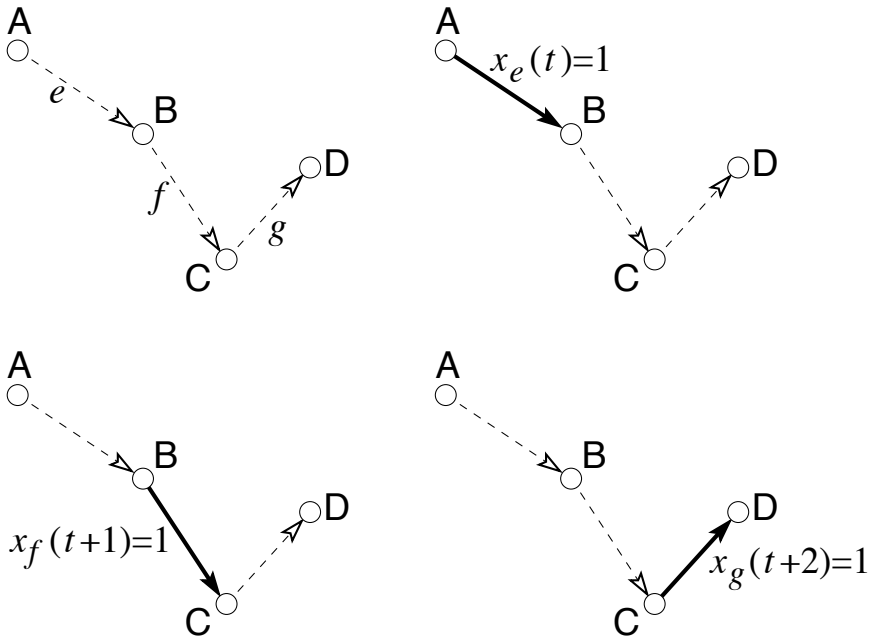



**Fig. 8.1.** Example of route obtained through link activation.

fixed nodes, this is not much of an issue for WMNs. Moreover, finding an efficient STDMA link activation pattern has the drawback of being computationally expensive. However, this can be done by a centralized unit (e.g., located in one of the MAPs, which are usually the most computationally capable among the MRs), which determines a proper transmission schedule and communicates it to the other nodes. This can be realized by broadcast messages or by piggy-backing this information in other control messages.

In the following, we will specify modalities according to which the 0–1 decision variables corresponding to link activation can be determined. In particular, it is not restrictive to focus on an *uplink* problem, i.e., on how to deliver a given amount of packets, known a priori, *from* all MRs *to* any of the MAPs in the shortest time. This problem can be generalized to a downlink problem (i.e., to activate links so as to deliver traffic from any of the gateways to all MRs), which is conceptually identical. In fact, the downlink problem can be solved by looking at an equivalent uplink problem with reversed delivery requirements (i.e., where packets are to be sent from nodes to gateways instead of the opposite). For the uplink problem, since we assume a directed graph, we should also reverse flow directions and link parameters (e.g., $g_{ij}$ must be changed into $g_{ji}$). The link activation pattern found for the uplink problem can be flipped over time to obtain the solution to the downlink problem.

However, activation variables $x_e(t)$ can not take arbitrary values. The management of link activations should satisfy feasibility conditions related to the physical nature of the problem. Among the key points which will be discussed in the follow-

ing, we highlight here the interference requirements, which forbid certain links from being simultaneously activated, since some of the resulting transmissions will not be successful. Considerations about interference are also often coupled with MAC protocol issues. In fact, as shown in many contributions (e.g., in [14]), in a centralized environment, deterministic access provided by an STDMA scheme obtains better performance than random access schemes such as the Distributed Coordination Function (DCF) of the IEEE 802.11 MAC. However, our STDMA scheduling might simply be a deterministic link activation pattern superimposed onto an underlying MAC protocol, which is designed for distributed and random access. Indeed, several contributions [6, 12, 15–17] make (explicitly or implicitly) this assumption and for this reason combine interference and MAC protocol issues when determining the compatibility of simultaneous link activation.

The most widely used classification of interference models in the literature dates back to [18] and distinguishes between the so-called *physical* and *protocol* interference models. In the former, the feasibility of simultaneous link activations is determined by the Signal-to-Interference-Ratio (SIR) of all receivers being above a given threshold. The latter imposes instead simpler interference conditions modeled through graph neighborhood relationships. Actually, more than of a single model, we should speak of *protocol models*. In fact, the protocol model was originally intended to represent the IEEE 802.11 MAC protocol (hence the name), but in some works a slightly different implementation can be found, especially when IEEE 802.16 is used instead, even though the interference model is still called the same. These issues will be discussed in detail in the following, and we will present a classification which also aims at solving some terminology inconsistencies.

In addition to these two classes there is another possible approach, i.e., to directly estimate the interference, e.g., by measuring it in the scenario of interest [19], or through higher layer statistics [20]. Since we take an *a priori* approach to interference characterization, we will not discuss this *measurement-based* interference model further. However, it is worth mentioning as the one which is, in a sense, adopted by some related contributions, especially those dealing with routing metrics, e.g., [20, 21].

## 8.3 State of the Art

There is a vast literature in the field of wireless networks. The increasing interest for WMNs has recently brought researchers to revise typical issues of wireless networks in the context of this emerging technology. Specifically, traditional research topics such as link scheduling, routing, channel assignment and topology control find in WMNs new challenges and applications, as WMNs raise challenges and problems which need new solutions as the existing ones do not apply directly.

In this section, we provide an exhaustive up-to-date review of the literature on routing, scheduling and related cross-layer approaches for WMNs. The research works are classified according to the investigated research topics so as to guide the

**Table 8.1.** Taxonomy of related work.

| Reference | Schd | Rout | ChAs | Interf. | Approach |
|---|---|---|---|---|---|
| Alicherry *et al.* [6] | ✓ | ✓ | ✓ | P | O,A |
| Tang *et al.* [7] | | ✓ | ✓ | P | O |
| Cruz, Santhanam [8] | ✓ | ✓ | | Φ | O |
| Brar *et al.* [10] | ✓ | | | Φ | A |
| Kodialam, Nandagopal (1) [12] | ✓ | ✓ | | | T,O,A |
| Kodialam, Nandagopal (2) [13] | ✓ | ✓ | ✓ | P | T,O,A |
| Jun, Sichitiu [15] | ✓ | | | P | T |
| Ben Salem, Hubaux [17] | ✓ | | | P | T,A |
| Draves *et al.* [20] | | ✓ | | M | O |
| Yang *et al.* [21] | | ✓ | ✓ | M | T |
| Wu *et al.* [22] | | ✓ | ✓ | M | A |
| Salonidis, Tassiulas [23] | ✓ | | | P | A |
| Djukic, Valaee [24] | ✓ | | | P | A |
| Jain *et al.* [25] | ✓ | ✓ | | P | O |
| Cao *et al.* [26] | ✓ | ✓ | | | O |
| Wei *et al.* [27] | ✓ | ✓ | | P | O,A |
| Subramanian *et al.* [28] | ✓ | | | Φ | A |

Schd = scheduling, Rout = routing, ChAs = channel assignment

For "Interf."(=interference): P=protocol model, Φ=physical model, M=measurement.

For "Approach": O=optimization framework, A=practical algorithm, T=theoretical results.

reader to the contributions of interest. In Table 8.1, we report a taxonomy of the reviewed works. For each work, the table indicates the research issues addressed, the assumption about the interference model and the proposed approach.

The rationale for TDMA scheduling over WMNs can be derived from the very general approach for multi-hop wireless networks presented in [11]. In [23], the authors proposed a distributed implementation of such an approach for ad hoc networks. However, the resulting rationale can be applied, with minor modifications, to WMNs as well. In this work, a fluid model is proposed to quantify link activations, and the resulting evaluations are used by the terminals so as to share the medium in a fair and entirely distributed manner. Wireless interference is characterized through the protocol model. Another related approach to address TDMA scheduling for a WMN was presented in [24], where again the protocol model was used.

This same interference model was also used in two different works, [15] and [17], where wireless mesh scheduling was investigated from a theoretical point of view. Among the contributions presented in these works, we highlight in particular that the former gives a lower bound on the length of the optimal WMN link activation pattern, whereas the latter determines an upper bound on the same value, and proposes a fair scheduling mechanism. In the following sections, we will revisit the theoretical results of these works and extend them so that they can be applied in a more general way, i.e., with any interference model.

Another work dealing with scheduling in WMNs is [26], where the specific case of IEEE 802.16 mesh mode operating with centralized scheduling was addressed.

Here, an optimization framework was presented to maximize system throughput under specific fairness constraints. However, no interference model was presented, since the link allocation is only limited by what will be referred to in the following as half-duplex constraint. A further point of interest of this work is that the authors considered a Pareto dominance approach to compare scheduling solutions and find the optimal one.

The limitations imposed by oversimplified interference models heavily affect scheduling, as shown in [10]. The main contribution of this work is to show that assuming the protocol interference model may lead to inefficiencies in the scheduler implementation, whereas taking the physical interference model into account can achieve better network performance. To this end, a fast heuristic algorithm was proposed which assumes pre-determined traffic weights on each link (which, e.g., can come from a routing algorithm executed a priori).

Like scheduling, routing is also a challenging task in WMNs. In this scenario, several works investigated the task of properly defining metrics to be used in routing algorithms [20–22, 28]. In [20], the authors introduced a routing metric which is computed through the estimation of the interference of the links belonging to a path by means of delay probes. This approach was extended in [21] and [22] to the multi-channel case by also including the channel assignment problem. The former work used a theoretical approach, whereas the latter presented a practical algorithm supported by experimental results. Finally, [28] included interference awareness considerations in the computation of the routing metric, by utilizing the physical interference model.

Topology control considerations were included in the routing investigation performed in [7]. In this paper, optimality conditions to derive routing under QoS constraints were studied for a multiple channel network. The protocol interference model was used.

In general, standard solutions based on shortest-path algorithms are very likely not to be suitable for WMNs [21]. In fact, routing metrics based on the minimum hop count may have poor performance because they try to exploit wireless links between distant nodes. These long wireless links can be slow and lossy, leading to poor throughput. Furthermore, the objective of a traditional shortest-path routing algorithm is usually in contrast with that of link scheduling algorithms. Assuming a predefined path between a source and a destination implies that any link scheduling algorithm is forced to activate only the links belonging to that path. The link scheduling may be sub-optimal in the sense that any scheduling algorithm is prevented from optimizing the exploitation of the available network resources.

A pipelined approach which addresses both routing and scheduling has been considered in some recent works. In [27], scheduling and routing are performed. The scenario is specifically an IEEE 802.16 mesh operating with centralized scheduling. Here, a two-step procedure was proposed. First, a route selection algorithm identifies low interference paths toward the destination. Then, scheduling is performed among the routes by considering compatible link activations according to the protocol interference model.

However, as shown in [25], scheduling and routing algorithms impact each other and their optimality is strongly coupled. In particular, after a review of interference models, [25] addressed the question of combining optimal link scheduling with sub-optimal routing and vice versa. The main conclusion is that interference-awareness is also beneficial at the routing level. In a more general sense, this also implies that a joint optimization of routing and scheduling [11] is the most preferable solution.

A framework for joint scheduling and routing was described in [12], where the authors introduced a heuristic technique to solve the joint routing/scheduling problem. Specifically, routing and scheduling were solved as optimization problems over an undirected graph. The authors considered communication links as compatible if they respect what we call duplex constraints, i.e., the number of transmissions and receptions that nodes can simultaneously perform are limited. No additional interference constraint was considered. The necessary and the sufficient conditions were then derived to guarantee the link scheduling feasibility. For a given pair of nodes the objective was to determine the maximum achievable flow rate under the duplex constraints and the link scheduling feasibility conditions. This can be formulated as a linear programming problem. The proposed solution ensures that the link scheduling is feasible as the scheduling constraints are considered when solving the routing problem. The scheduling of each flow was then performed by coloring the network graph with a known graph coloring algorithm [29]. In [13], the authors extended their model to multiple channels and the protocol interference model. Specifically, they derive both necessary and sufficient conditions for a feasible channel assignment and scheduling in a multi-radio network. Again, the channel assignment problem was modeled as a linear optimization problem. Additionally, a heuristic algorithm is proposed for solving the problem. A similar approach was proposed in [6]. In this paper, the authors mathematically formulated a joint channel assignment and routing framework, taking into account the protocol model interference constraints, the number of channels in the network, and the number of radios available at each mesh router. Within this framework, they devised a heuristic to perform routing and channel assignment aimed at optimizing the network throughput performance.

Finally, in [8] a joint analysis of routing and scheduling for multi-hop networks was presented, which also includes power control. Another interesting aspect of this paper is that it addresses half-duplex limitations of the wireless medium, as well as directionality of links and the physical interference model. However, the paper does not directly investigate WMNs, but rather it mainly focuses on systems similar to ad hoc or sensor networks, since the objective of the optimization is the minimization of the power consumption, which is not an issue in WMNs.

To sum up, joint and cross-layer approaches were proposed by several papers dealing with routing and scheduling for WMNs, but the formulation of an overall framework which encompasses all of these issues is still an open field of research. Existing approaches are often unsuitable for WMNs due to dissimilar optimization goals and/or oversimplified interference models. The formulation of a comprehensive framework for these issues, also addressing technological issues in a realistic manner and correctly taking into account link directionality, duplex constraints and different possibilities for the inter-link interference model, is a promising goal for

future research. As a first step in this direction, we will give in the following some guidelines and analytical insights on the performance of joint routing and scheduling in WMNs.

## 8.4 Problem Statement

To study routing and scheduling under the graph formulation reported in Section 8.2, we will use the language of constrained linear programming problems, as this is an approach commonly used to decide the assignment of $x_e$ variables [6, 12]. We will therefore speak of *constraints* to describe any limitation imposed to the activation of links by MAC and physical layers.

These constraints can be of different natures, and we will describe them in separate subsections. First of all, edge activation implies node activation for transmission and reception, for which there are limitations on the transceiver capabilities of each node involved. As will be discussed in Subsection 8.4.2, the activation of edge $(i, j)$, which employs $i$ and $j$ as transmitter and receiver, respectively, may not be feasible, if these nodes participate in other link activations.

Moreover, links which involve different nodes for what concerns both the transmitter and the receiver, might or might not activate simultaneously depending on the mutual electro-magnetic interference. For this reason, we need to define a compatibility relationship among the links in the network. Several models for this will be reviewed in Subsection 8.5.3. In most cases, they can be subdivided into the two main classes of *protocol* and *physical* model, already mentioned in Section 8.2. To better understand the "protocol model", in Subsection 8.5.3 we will also briefly discuss the underlying assumptions of IEEE 802.11 and IEEE 802.16 standards for what concerns access control.

Prior to investigating in detail these constraints, we give an overview of other related problems which can be framed within our approach, which is the goal of Subsection 8.4.1.

### 8.4.1 Channel Assignment and Node Placement Framed into the Model

Some issues discussed previously in Section 8.3 can be incorporated in our framework. For instance, it is common to assume that the wireless medium has several channels available for transmission. From a simplified point of view, these channels are often considered orthogonal [12, 22, 30] and it is further assumed that the MRs own different NICs so that they can communicate on many channels in parallel. A relevant point in this case is whether the terminals can rapidly change the channels on which their NICs are active. With current state-of-the-art technology [31], the order of magnitude of channel switching time can be $0.1$ s, which is likely to be much higher than one time-slot; thus, we need to assume that the assignment is not modified during the schedule, and every node can be active only on certain channels of choice.

For this reason, the study of channel assignment in this case mostly relates to routing, and corresponds to identifying low-interference paths whose parallelism is further improved by the presence of orthogonal channels. In fact, links which would interfere if scheduled jointly can be activated together if their transmitters and receivers are tuned to different channels. The issue of channel assignment in the orthogonal case is therefore often seen as a graph-coloring problem, where colors assigned to edges represent orthogonal wireless channels. From the perspective of our link activation framework, the orthogonal multiple-channel assignment can be incorporated following a similar rationale, e.g., by defining variables $x_e^{(c)}(t)$, where the additional color index $c$ spans over a set of channels $\mathcal{C}$ and denotes the channel possibly used by link $e$. This imposes additional constraints, i.e., that the number of activated channels for a node is less than or equal to a given parameter, corresponding to the number of NICs it owns, and that a link can be activated only if transmitter and receiver share a common active channel.

In general, the additional challenges imposed by the presence of multiple orthogonal channels are not further considered here, since they are out of scope of our analysis. Note only that, from a purely mathematical point of view, if frequency is considered as a perfectly separable resource, differences between frequency-division and time-division multiplexing are limited and they can be translated into our framework. For this reason, most of the conclusions we will draw in the time domain also hold for orthogonal multiple channel assignment.

An important observation raised, e.g., in [32], stems from the observation that, in real network systems, contiguous channels are not perfectly separated at the physical level, but are instead partially overlapping. In general, this is regarded as an undesired effect and to deal with it channels are assumed to have guard bands that are not used for transmission. It is, for example, usual to limit the use of IEEE 802.11 MAC to channels 1, 6, and 11, which can be considered as orthogonal with a good degree of accuracy, leaving the remaining channels unused [16]. However, an entirely different approach was used in [32] and related papers. These contributions show that the existence of partially overlapping channels, instead of being a problem, may turn into an advantage for the network if properly exploited. In particular, it is possible to partially obtain transmission parallelism even by using a single NIC. Intuitively speaking, this happens as the intended transmitter and receiver do not need to be tuned to the same channel, but they can choose two different partially overlapping channels. The choice of the channel to which a node tunes is therefore a trade-off between maximizing the overlap for useful connections and minimizing it for the interference it causes to other links when transmitting.

Such an extension to multiple overlapping channels can be framed in our joint routing and scheduling framework, even though it would require a long analysis which can not be reported here for space reasons. However, we consider it as a possible interesting subject for future research.

Another possible related investigation is the evaluation of the network deployment, especially for what concerns node placement (MRs and MAPs). In most of the related work it is assumed that the nodes' positions are decided *a priori*. The reason for this is twofold: on the one hand, it is realistic to think of network deployment as

realized in a different design phase than routing and scheduling; on the other hand, it is also difficult to allow for an entirely free node placement, due to physical and environmental constraints, as well as the not-in-my-backyard problem. Nevertheless, it is still possible to allow a certain degree of choice without violating realism. This can be done by following the approach presented in [33], and adapted to WMNs in [34]. The problem statement is slightly changed, so that nodes of the graph no longer represent terminals but are instead *candidate positions* where terminals can be placed (in [33] terminals are UMTS base stations, whereas in the WMN case, they are MRs and MAPs). An additional binary decision variable $y_n$ is introduced for every $n \in \mathcal{N}$ to denote whether position $n$ is actually occupied by a terminal or not. The rest of the analysis proceeds identically, with the only modification of requiring any edge activation variable $x_{(i,j)}$ to be less than or equal to both $y_i$ and $y_j$, as a communication link can be actually activated only if both its ends correspond to physically deployed terminals.

### 8.4.2 Transceiver Constraints

Our graph-based approach determines a joint scheduling and routing through link activation. As communication links are represented through edges of the graph, most of the constraints are edge-based, i.e., they must be respected by every active edge. However, the first important constraint we discuss is node-based, i.e., it has to be evaluated at every node, and relates to the fact that the node capabilities for transmission and reception are limited. In particular, we focus here on narrowband channels, where it is not possible to receive simultaneously from multiple sources. We remark that special techniques, such as Wideband Code-Division Multiple Access (WCDMA) [35] or Multiple Input Multiple Output (MIMO) [36] channels, can improve this condition. However, they are out of the scope of our investigations. In the following, we therefore assume that at most one signal can be decoded, and any other transmission the receiver is able to listen to can only be regarded as interference. The presence of interference at the receiver does not necessarily mean that the packet can not be correctly decoded. As will be shown in the next subsection, the interference model comes into play at this point. If the protocol model is used, any superposition of signals will result in a collision, i.e., no packet can be received. In the physical model, the strongest received signal may still be successfully decoded. However, regardless of the interference model, the maximum number of possible simultaneous successful receptions is *one*.

A similar situation happens for the transmitter. Even though on the wireless medium it is possible to operate in a multicast fashion, i.e., from one transmitter to many receivers, in this case the *same* transmission takes place for all of them. Note also that multicast transmission, which would require additional specifications, e.g., for duplicated packet control, does not correspond to the problem we consider, where the intended destination is only one. For these reasons, we will assume in the following that multiple transmissions from the same node are forbidden. However, we remark that the issue of exploiting the possibility for some relay nodes to listen to the communication, even when they are not the intended receivers, to improve the

network connectivity by exploiting cooperation [37] or network coding [38] is a very promising subject for future research in wireless networks.

Finally, not only can simultaneous transmissions and receptions be at most one, but also the wireless communication medium is intrinsically *half-duplex*, i.e., a node can not listen on the same channel on which it is transmitting at the same time, or the transmitted signal will jam any packet reception [39]. Possible solutions to this problem, so as to realize a sort of full-duplex communication with simultaneous transmission and reception at a node, can be to utilize more than one NIC to exploit the possible presence of multiple channels [16, 22], or to use multiple directional antennas [40, 41]. However, these techniques do not entirely solve the problem, as they obtain full-duplex capability at the price of additional resources. Moreover, they decrease network connectivity, which in certain cases can be an undesirable effect, as the nodes should use compatible channels or antenna beams. Finally, we remark that in the multiple channel case any NIC is still utilized in a half-duplex fashion, i.e., no simultaneous transmission and reception is still possible on the same channel. For these reasons, we impose that the activation of links should satisfy the constraint of not activating more than one operation (i.e., either a transmission or a reception) for each node. Formally, this constraint translates into the following:

$$\forall i \in \mathcal{N}, \forall t: \qquad \sum_{j \in \mathcal{S}_i} x_{ji}(t) + \sum_{j \in \mathcal{R}_i} x_{ij}(t) \leq 1. \tag{8.3}$$

Apparently, the importance of including the half-duplex aspect in this constraint is often underestimated when modeling multi-hop wireless networks. In fact, the need for such a constraint is rarely mentioned. This may be due to the fact that, as already emphasized, most of the investigations use the protocol interference model which, as discussed in the following, prevents simultaneous transmission and reception at the same node from happening. However, we believe that it is important to distinguish the edge-based interference constraints from the node-based duplexing limitation. Indeed, the interference constraint does not necessarily translate into the protocol model, which can be replaced, e.g., by the physical model. Instead, the duplexing limitation holds irrespective of the interference model. For this reason, we will always impose the half-duplex constraint in any problem formulation. Note also that our assumption in this respect might seem different from [12], where the authors allow for the possibility of using both directions of the link at the same time in what they call *full-duplex* case. However, this case is used in conjunction with the protocol interference model. In general, any case where full-duplex nodes are mentioned does not refer in reality to the possibility of transmitting and receiving on the same frequency *at the same time instant*.

Apart from this very general constraint, other limitations to the simultaneous activation of edges make specific assumptions on the nature of radio interference and on the underlying MAC protocols. We will review both these aspects in the following section.

## 8.5 Interference Models and Relationships with MAC Protocols

In Section 8.2 we mentioned the need for using uni-directional edges in our network graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. First of all, this subsection aims at motivating this choice in more detail. After this explanation, we outline some aspects of well-known MAC protocols and finally we review which model of mutual interference among nodes can be used to determine if the activation allows the correct reception of all transmitted packets.

### 8.5.1 Link Directionality

The choice of using the directed graph representation, which captures the *anisotropic* nature of wireless links is surely more realistic from the physical point of view. In fact, wireless links are characterized by strong asymmetry [42]. Due to environmental limitations and also different power levels, it is even possible that two nodes $i$ and $j$ belonging to $\mathcal{N}$ are linked only one-way, i.e., $(i, j) \in \mathcal{E}$ but $(j, i) \notin \mathcal{E}$.

However, our choice is not only motivated by the desire to better adhere to reality. In fact, the frequent assumption of bi-directional links is in most cases due to the application of the analysis to IEEE 802.11 scenarios. As the IEEE 802.11 standard is supposed to work on entirely reliable links only, edges in $\mathcal{E}$ need to be bi-directional, This also relates to the choice, which will be discussed in the next subsection, of modeling interference with the protocol interference model, in its implementation more closely related to IEEE 802.11.

Yet, the decrease in the problem complexity gained with bi-directionality assumption is marginal (the number of edges is only decreased by a constant factor of 2), and implies an oversimplification in modeling interference conflicts [9], especially when focusing on a centralized STDMA scheme, if an underlying IEEE 802.11 MAC is not employed. Instead, not only is the problem version with directed edges of $\mathcal{E}$ more accurate, but it also includes the undirected graph as a special case.

### 8.5.2 Overview of MAC Protocols

To realize distributed medium access with low cost technology, random MAC protocols are often used. In particular, the IEEE 802.11 standard has obtained a great success for what concerns its DCF-based version, operating with *four way handshake*, which implies that the transmission is initiated after a successful request-to-send (RTS) and clear-to-send (CTS) exchange, and after the data transmission an acknowledgement (ACK) is also to be sent from the receiver to the transmitter.

However, IEEE 802.11 is known to suffer from many problems, which are severely limiting for WMNs. In fact, to operate in a totally distributed manner, IEEE 802.11 requires the transmission of many control packets, whose overhead is often heavy. In a WMN most of them are not necessary since most of the control can be centralized. Moreover, its collision avoidance mechanism often imposes unnecessary constraints which limit network parallelism, especially because it does not properly capture wireless interference. Finally, the main advantage of the conceptual simplicity and ease of implementation of the IEEE 802.11 MAC is not strictly required in

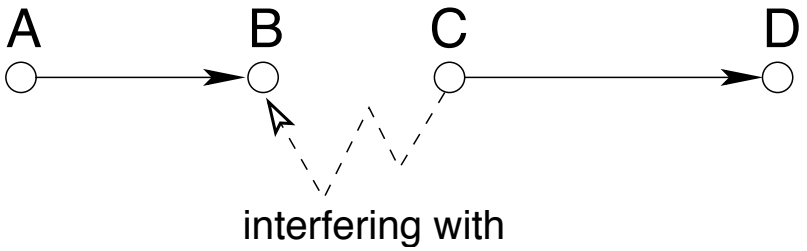



**Fig. 8.2.** A case of transmission showing hidden terminal problem.

the WMN backbone, which is composed of more expensive and technologically advanced terminals. Anyway, note that IEEE 802.11 can still be used to interface a MR with its MCs; however, as has been already said, this part of the network is not investigated in our analysis.

Another standard which is envisioned to have applicability for WMN is IEEE 802.16 [4], which besides the point-to-multipoint (PMP) mode is also available in a mesh mode. We focus on the following distributed scheduling version. IEEE 802.16 aims at partially solving some of the aforementioned problems as, unlike IEEE 802.11, it utilizes a random access-based procedure in the control frame with a *three-way handshake* procedure, where a Request is answered by a Grant, which is finally followed by a Confirm message from the transmitter. Part of the advantage of IEEE 802.16 stems from the additional requirement for topology awareness, which is exploited in the *distributed election* mechanism to guarantee that no collision arises in the control message exchange.

The reason for these protocols to include specific handshaking procedures, and possibly also further random decisions and exponential backoff algorithms, is to cope with the fact that nodes can operate in a distributed fashion. In fact, random medium access protocols in IEEE 802.11 and IEEE 802.16 potentially suffer from problems due to uncoordinated transmissions. One well known inefficiency of random access protocols is the *hidden terminal problem*. This occurs when a node transmits, being unaware of other ongoing transmissions, which will cause a collision at some receivers.

An instance of this problem is shown in Fig. 8.2, where terminal A transmits to B and C transmits to D. Assume that A and C can both transmit to B but not to each other; hence, the reference to C being "hidden" to A and vice versa. In this case, A is unaware of C's transmission, which can be harmful for the reception at terminal B. Conversely, also node C is not informed of A's intention to transmit to B. Thus, a collision will occur at B, i.e., presence of strong interference, which is generally assumed to cause inability for the receiver (in this case, node B) to successfully decode the packet.

It is very easy to construct other similar examples of hidden terminal problems, see also [43, 44] where the interested reader can find further details. In general, the hidden terminal problem affects the transmission efficiency in the sense of causing possibly erroneous transmissions, which result in wasted bandwidth.

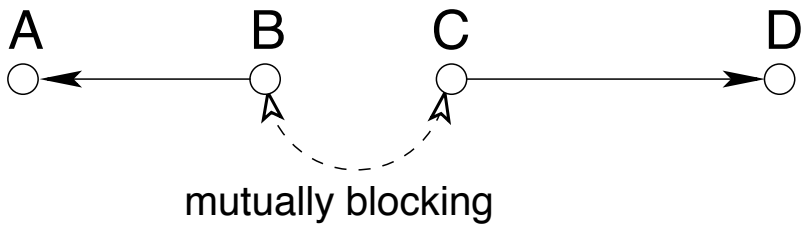



**Fig. 8.3.** A case of transmission showing exposed terminal problem.

At the same time, a similar issue with different consequences is the *exposed terminal problem*, which is exemplified in Fig. 8.3. Here, B and C intend to transmit to A and D, respectively. This time, the wireless medium is used inefficiently as both transmissions could be accomplished in parallel, but the senders are instead "exposed" to each other; thus, one of them transmits but the other refrains from sending packets as soon as it listens to the transmission of the other node, considering that it could cause collision. In this case, the medium access procedure is inefficient due to the low channel utilization not fully exploiting both possibilities of sending data over the channel.

Actually, both IEEE 802.11 and IEEE 802.16 MAC protocols aim at partially solving these problems. The *four-way handshake* mechanism of DCF tries to avoid the hidden terminal problem, since, e.g., a CTS sent by the intended receiver silences other potential transmitters which listen to it, thus blocking their transmissions. However, the exposed terminal problem is still unsolved, and is often considered as one of the main reasons of IEEE 802.11's inefficiencies [14]. The mesh mode of the IEEE 802.16 standard operates similarly to avoid the hidden terminal problem, since the three way handshake can work in the same way. Moreover, the distributed election mechanism allows to alleviate the exposed terminal problem.

However, if applicable, a perfectly centralized medium access, which follows a pre-determined collision-free schedule, would work even better to prevent such situations from arising. A WMN would be theoretically able to apply a centralized STDMA schedule with the only constraints of the half-duplex limitation and the physical interference (i.e., without additional limitations to the transmission parallelism imposed by random MAC protocols), which clearly obtains better performance than more constrained cases. In general, centralized control can have disadvantages due to delay in collecting the information from the whole network, which causes the topology awareness to be inaccurate. However, in a WMN MRs are fixed and the network topology is therefore relatively stable, thus this solution is likely to be preferable to distributed algorithms. Still, centralized scheduling is also applicable if an underlying MAC protocol (in particular, either IEEE 802.11 or IEEE 802.16) is present, even though the performance will be suboptimal due to the additional protocol constraints.

As a side comment, note also that, in spite of the aforementioned techniques to solve them, the hidden/exposed terminal problems may be present in case of link

asymmetry and/or time-varying channel. This happens because the rationale behind these protocols assumes that hidden transmitters are necessarily in the reception neighborhood of the potential receiver. However, this is true only if $g_{ij} = g_{ji}$. If this condition is not verified, a node can be unaware of hidden terminals even after a successful handshake exchange. Similarly, in the asymmetric channel condition, some nodes can become aware that they are exposed terminals only when certain interfering nodes, which are unable to listen to the packets of those nodes, start transmitting. Finally, due to erratic behavior of the wireless channel, it might happen that topology information collected at a single node is outdated or wrong. Also in this case, centralized control would help the network management in identifying and solving inconsistent information, whereas if the nodes operate in a distributed fashion, the effect of the hidden or exposed terminal problem may be stronger.

### 8.5.3 Characterizing Interference

The contribution in [18], besides having settled the basis for information-theoretic studies on the capacity of wireless networks, also introduced two useful models of interference among radio transmissions. Following their classification, we refer to them as *protocol* and *physical interference model*, respectively. Indeed, the literature reports several variations of these models, which we review below. For simplicity, we will avoid more complicated extensions which model transmission aspects such as directional antennas, capture effect (when modeled with a threshold) and so on. An overview about this can be found in [45].

#### Protocol Interference Models

The protocol interference model, in its original version, follows the rationale behind the IEEE 802.11 MAC. It models interference as causing *collision*, i.e., impossibility of correctly decoding a received packet, if other nodes in the network simultaneously exchange messages with sufficient power to disturb the ongoing transmission. The main advantage of an interference description through the protocol model is its conceptual simplicity, and the ease of mathematically formalizing the resulting interference conditions. We believe that this is, in fact, the main reason for the widespread use of the model.

The rules of the protocol interference model simply forbid that certain transmissions are simultaneously activated, when it is assumed that they will cause collision. It should be noted that, in spite of the node-based nature of the interference, this criterion is modeled through an *edge-based* constraint, i.e., to be verified for any active edge. As reported in [6, 12], a way to formalize this constraint is to define a conflicting set of edges $\mathcal{I}(e)$ associated to any edge $e \in \mathcal{E}$. According to the notation employed, the set $\mathcal{I}(e)$ may or may not include $e$ itself. In the following, we will tacitly assume that $e$ is included in $\mathcal{I}(e)$. The required condition is then that if edge $e$ is active, its associated set $\mathcal{I}(e)$ must contain no more than one active edge (i.e., $e$ itself). Formally,

$$\sum_{f \in \mathcal{I}(e)} x_f(t) \leq 1 \quad \text{if link } e \text{ is active at time } t, \text{ i.e., } x_e(t) = 1. \tag{8.4}$$

In other formulations where $e$ does not belong to $\mathcal{I}(e)$, the condition above can be promptly modified by imposing the sum of activity variables over $\mathcal{I}(e)$ to be 0 if $e$ is an active link.

Sometimes, this relationship is translated into a *conflict graph* $\mathcal{G}_C = (\mathcal{E}, \mathcal{L}_C)$ where conflict relationships among edges are represented [16, 17]. In this formulation, the *nodes* of graph $\mathcal{G}_C$ are the *edges* of $\mathcal{G}$, whereas every edge of $\mathcal{L}_C$, which is a pair $(e, f)$ with $e, f \in \mathcal{E}$, represents that $e \in \mathcal{I}(f)$. Though conceptually nice, this representation turns out to be very impractical in most cases, since $\mathcal{E}$ usually contains many more elements than $\mathcal{N}$ (in the worst case, $|\mathcal{E}| = |\mathcal{N}| \cdot (|\mathcal{N}| - 1)$, where $|\cdot|$ is the cardinality of the set). Also, considering the conflict graph does not solve the problem of the high computational complexity of graph operations (usually NP-complete), rather sometimes it worsens it, due to a larger graph size. Finally, we remark that the contributions which utilize the conflict graph representation consider the WMN to be a bi-directional graph, thus they utilize a bi-directional version of the conflict graph. However, this worsens the problems with respect to the link directionality issue. In fact, in this way it is impossible to describe that $e$ causes a collision at $f$ but not vice versa. For these reasons, the conflict graph description, introduced here for the sake of completeness, will not be mentioned further in our analysis, but the simpler approach based on the set of conflicting edges $\mathcal{I}(e)$ will be used.

The way to determine this set depends on which MAC protocol is used. In the literature, there are subtle differences among its definition, since some authors refer to the protocol model albeit they have in mind a different access strategy than IEEE 802.11 MAC or implicitly implement some protocol improvement. We refer to them as the *class of protocol interference models*, which actually encompasses several mathematical formulations. In the following we will speak of protocol model without any further specification only when describing general properties of the class. Otherwise, a specific version of the model will be mentioned.

Before describing other more complicated versions, we intentionally introduce a very simple model belonging to the protocol interference class. One straightforward possibility of defining $\mathcal{I}(e)$, though also an extreme one, is to consider $\mathcal{I}(e) = \mathcal{E}$ for all $e \in \mathcal{E}$, i.e., at most one edge can be activated at any given time throughout the whole network. In other words, either exactly one edge is active, or no edge is active at all. Due to this property, we refer to this version as the *01protocol* model. Even though it is quite oversimplified, it can be useful as a theoretical term of comparison. In fact, the *01protocol* model is clearly the worst possible case of interference condition, where space diversity can not be exploited to obtain transmission parallelism.

Actually, this situation necessarily occurs on certain special topologies. For instance, in [45] this model is mentioned as used in [46] to derive the performance of DCF in an IEEE 802.11 hot-spot controlled by a single access point. Indeed, it is true that the *01protocol* model holds here, but the reason is not electromagnetic interference, but rather that the topology is a star network (every node is connected only to the access-point). Therefore, the reason for having such a constraint of at most

one link activation at any given time stems from the transceiver constraints, not from interference. We emphasize that this limitation should not be confused with interference constraints. Apart from their different motivation, already discussed in Subsection 8.4.2, the *01protocol* is clearly a more restrictive condition than the transceiver constraint (i.e., the interference constraint described by the *01protocol* is a sufficient condition for the duplex constraint). It may happen that the transmission parallelism is very difficult to obtain due to physical reasons. In certain cases, the propagation environment may exhibit extremely low attenuation from path loss so that interfering signals propagate for very long distances. If this is the situation, the *01protocol* model can be appropriate to capture such weakness of the links even if the topology is loosely connected. On the other hand, as has already been mentioned the duplex constraint holds true for any single-channel network regardless of the interference model.

Apart from the simple *01protocol* model, other versions need to rely on propagation aspects, though still simplified, to be formally described. It is common in the literature [16, 17, 44] to adopt a simple approach which makes use of geometric considerations, by implicitly assuming omni-directional propagation, isotropic environment and absence of fading. Actually, these assumptions are introduced only for the sake of presentation, as they are clearly unrealistic from the transmission physics point of view. Note however that it is possible to remove them without changing the rationale.

First of all, define the concepts of *coverage* and *disturbance* of a node.[2] Node $i$ is said to *cover* node $j$ if a transmission from $i$ can be correctly received by $j$ in the absence of any other transmission (i.e., the only factor degrading the signal quality is the thermal noise at the receiver). This means that an edge $(i, j)$ exists in $\mathcal{E}$, and therefore, $j \in \mathcal{R}_i$, or identically $i \in \mathcal{S}_j$. Similarly, node $i$ is said to *disturb* node $j$ if $j$ can detect that $i$ is transmitting, even though it may not be able to decode the message. The coverage relation is clearly a sufficient condition for disturbance, but not necessary. It is also common to find these relationships as translated into the definition of a *coverage area* and a *disturbance area*.

Following the line of neglecting several propagation effects and considering only the distance-based path loss, a so-called *transmission range* can be defined. Note that, in the literature on ad hoc networks, this range is often considered equal for all nodes. For WMNs this might be a strong approximation, since nodes may be considerably heterogeneous. Moreover, another weak point of this definition is that the distance up to which a communication link can be activated does not depend on the transmitter's characteristics (in particular, on its transmitted power) only, but also on the receiver's sensitivity. However, the transmission range assumption can be relaxed without changing the rationale, so we leave them only for presentation reasons. Thus, in the following we assume that coverage and disturbance areas are circular with radius equal to the transmission range and to a given constant $\vartheta$ (usually larger than 1)

[2] The term which is most widely used [6, 14] for the latter is "interference." We use the term "disturbance" to avoid confusion for the reader, as the term "interference" is used in our analysis with a broader meaning, and does not necessarily refer to the protocol model.

times the transmission range, respectively. Formally, if $r_{\mathrm{TX}}$ is the transmission range, the coverage and disturbance area for a node $n$ are two-dimensional balls centered on $n$ (i.e., on its location) with radii $r_{\mathrm{TX}}$ and $\vartheta \cdot r_{\mathrm{TX}}$, respectively.

In the original and more common version, which we call hereafter *11protocol model*, it is implicitly assumed that the IEEE 802.11 MAC is employed. For this reason, we make the assumption that links are bi-directional, as IEEE 802.11 is designed to work for bi-directional links only, and heavily relies on this hypothesis.

Following the IEEE 802.11 MAC, the *11protocol* model dictates that a transmission on $(i, j) \in \mathcal{E}$ is interference free, and can therefore be activated, only if there are no transmitters, nor receivers, belonging to any active link, with either $i$ or $j$ in the disturbance area, apart from $i$ and $j$ themselves. Remember that, to enable the transmission, the IEEE 802.11 MAC protocol requires that both $i$ and $j$ are in the coverage area, and thus also in the disturbance area, of each other.

Note that the reason for requiring the absence of interferers in both *receiver's and transmitter's* disturbance area of both interfering *transmitters and receivers* is that the IEEE 802.11 standard forces the receiver to acknowledge RTS and data packet with CTS and ACK, respectively. In other words, due to the four way handshake, a logical receiver is also a physical *transmitter*, therefore, it can cause disturbance to others. Similarly, the logical transmitter needs to perform *reception* (i.e., to receive CTS and ACK), for which it has to be collision-free. The four way handshake of IEEE 802.11 exactly aims at avoiding the hidden terminal problem on both forward and reverse link (the existence of which is specifically required by the protocol).

However, if IEEE 802.11 MAC is not used in the WMN backbone, there is no reason to impose such restrictive constraints, e.g., to mute a node which potentially disturbs the transmitter, but not the receiver. Note that to see this we need uni-directional edges, which we previously claimed to help in reducing unnecessary constraints on multiple transmissions, beyond being a better model *per se*.

In particular, these conditions can be relaxed if the IEEE 802.16 MAC is used instead. There are differences, not discussed here since they are out of scope of the analysis, between the four-way and three-way handshake, which do not only involve the packets exchanged, but also the aforementioned relationships of disturbance among nodes. We can then formulate a *16protocol interference model*, which proceeds identically to the *11protocol* model, with the notable exception that a collision is determined only when the designated *receiver* falls within the disturbance range of another *transmitter*. Any other combination (transmitter is under coverage of an interfering transmitter, or another receiver covers either the receiver or the transmitter) does not do any harm.

The *16protocol* model solves not only the hidden terminal, but also the exposed terminal problem, and it better accounts for the directionality of the wireless links. In particular, note that the condition of interference of the *16protocol* model refers to the intended receiver being under coverage of an interfering transmitter, not vice versa, since these conditions may not be equivalent.

A possible definition of $\mathcal{I}(e)$ in the *11protocol* model is thus

$$\mathcal{I}(e) = \{f \in \mathcal{E} : \text{transmitter or receiver of } f \text{ disturbs transmitter or receiver of } e\} \tag{8.5}$$

whereas in the *16protocol* model it is

$$\mathcal{I}(e) = \{f \in \mathcal{E} : \text{transmitter of } f \text{ disturbs receiver of } e\}. \tag{8.6}$$

Note that in both definitions $\mathcal{I}(e)$ includes $e$ itself.

In most of the works dealing with WMN backbone management, the *11protocol* model is what is meant when the protocol model is cited. However, if links are not bidirectional and the MAC does not follow the IEEE 802.11 standard, and especially if the IEEE 802.16 standard is used instead, there is no reason for using the *11protocol*, and the *16protocol* model would be more appropriate.

The general behavior of the model heavily depends on the ratio between the disturbance range and the coverage range. Apart from being in general hardware dependent, this value is also hard to quantify exactly, since the concepts of disturbance and coverage themselves have a vague physical meaning. In most cases, $\vartheta$ is arbitrarily chosen between 1 and 2, e.g., 1.6. This follows the approach commonly used, e.g., in sensor networks, where it is however conceptually more appropriate due to the fact that nodes are homogeneous (a condition which does not hold in WMNs).

If $\vartheta$ can be taken equal to 1, both *11protocol* and *16protocol* model can be translated to a simpler formulation connected with graph neighborhood relationships. In fact, in the case $\vartheta = 1$, the coverage range is equal to the disturbance range, and the coverage relationship (which is always necessary, but also sufficient for the disturbance if $\vartheta = 1$) is implicitly assumed in determining the existence of an edge in $\mathcal{E}$ between a transmitter and a covered receiver. Thus, node $i$ disturbs $j$ if and only if they are neighbors. The exact kind of neighborhood depends on which version of the protocol model is considered.

For the *11protocol* model,

$$\mathcal{I}((i,j)) = \{(k,\ell) \in \mathcal{E} : \{i,j\} \cap (\mathcal{R}_k \cup \mathcal{R}_\ell) \neq \emptyset\} \tag{8.7}$$

whereas for the 16protocol model

$$\mathcal{I}((i,j)) = \{(k,\ell) \in \mathcal{E} : j \in \mathcal{R}_k\}. \tag{8.8}$$

From this formulation, it is clear that the *16protocol* model simplifies the *11protocol* model as it considers the receiver $j$ being in the coverage range of an interfering transmitter $k$ as the situation where collision occurs. The *11protocol* model instead considers four possible combinations as colliding, i.e., all cases where $i$ or $j$ is under coverage of either an interfering transmitter $k$ or an interfering receiver $\ell$.

This last formulation of the protocol model through neighborhood relationships is very common in the literature. We briefly remark that it can be extended to cases where the disturbance area is larger than the coverage area, i.e., $\vartheta > 1$. This happens by considering an extended graph with virtual edges $\mathcal{E}_I$, which can not be activated as useful communication links but simply describe the interference relationships. The

one-hop output neighborhood $\mathcal{R}_i$ of a node $i$ can then be replaced by a larger set $\mathcal{R}'_i$ defined similarly to what reported in (8.1) but replacing $\mathcal{E}$ with $\mathcal{E} \cup \mathcal{E}_I$.

To sum up, the protocol interference model is easy to implement, and it offers several possibilities both to describe MAC aspects, which have been classified in the three different versions (*01protocol*, *11protocol*, *16protocol*), and to employ the preferred mathematical model (coverage/disturbance range, conflict graph, neighborhood relationships). However, these practical advantages come at the price of some theoretical drawbacks. In fact, all versions of the protocol model are imperfect in capturing wireless interference. First of all, the notion of coverage range (or equivalently, conflict graph or node neighborhood) is not entirely realistic. If several power levels are adopted, it is not possible to define a single measure of coverage even from an abstract perspective. Differently from, e.g., motes of a sensor network, the MRs may be heterogeneous devices, and therefore, may have unequal characteristics in terms of transmit power, receiver sensitivity, installation site and so on. Thus, it is very hard to summarize all these physical layer effects under a single item, e.g., a single coverage range.

Moreover, a definite reason of criticism against the protocol model is that interference is not a binary relationship [10, 25]. It is true that the *outcome* of interference evaluations can be reasonably limited to two values, i.e., the activation of multiple links is either interfered or interference-free. However, the *number* of involved nodes and edges, especially in large topologies, is larger than 2, and the "disturbance" relationship defined above does not correspond to a well stated binary operation when other communication links are active in the network.

For example, strong interference, which leads to packet loss, may be present in case three specific edges are simultaneously activated, but not when any two of them are. Thus, no specific link alone causes interference, but the problem is the joint effect of all links. Seen from the point of view of a single edge $e$, it might happen that $f, g \in \mathcal{E}$ can individually coexist with $e$, but not jointly. In this case, it is doubtful whether $f$ and $g$ should be inserted in $\mathcal{I}(e)$. We remark that usually the conflict set $\mathcal{I}(e)$ is evaluated pair-wise, as defined above, which would lead to problems as the joint activation of $f$ and $g$ is not prohibited. On the other hand, the alternative approach where any edge possibly disturbing $e$ (even if this happens only if other links are activated as well) is put into $\mathcal{I}(e)$, would be too conservative to be practically useful.

**Physical Interference Model**

These problems can be overcome by means of the physical interference model, whose rationale is as follows. The packet error rate (PER) at the receiver is a monotonically decreasing function of the Signal-to-Interference-and-Noise Ratio (SINR). It is often reasonable to simplify this relationship and consider a threshold approach, where it is assumed that a packet is correctly received with probability 1 if the SINR is above a given threshold. A way to formalize this is as follows:

$$\frac{P_i g_{ij}}{\sum_{k \neq i} P_k g_{kj} + N_j} \geq \gamma_j \tag{8.9}$$

where $(i, j)$ is the link of interest, the index $k$ in the lower sum denotes a possible interferer ($i$ is in fact excluded from the sum, as it is the intended transmitter), $P_x$ is the power emitted by node $x$, $g_{xy}$ is the path gain from $x$ to $y$ and $N_j$ is the noise at the receiver node $j$. The value $\gamma_j$, which defines the SINR threshold, can be in general a different value for every node $j$.

Hereafter we use these assumptions, which are made only for ease of exposition, but without loss of generality, as avoiding them would only lead to a more cumbersome (though conceptually identical) formulation. We take $\gamma_j = \gamma$ for all $j$. We also neglect the noise terms and we consider an equal power level $P$ among all transmitting nodes. In particular, the last assumption is equivalent to assuming that the power level is simply fixed. If this is the case, the elements $(g_{ij})$ of the matrix $\mathbf{G}$ can be replaced by $g'_{ij} = P_i g_{ij}$ and the power term can be omitted. If the power level is instead not fixed, it would become necessary to also include power control in the analysis. However, this can be performed within a very similar framework, as shown in [8].

In the context of our framework which describes scheduling and routing through link activation patterns, the constraint can be formalized as follows:

$$\frac{x_{ij}(t) g_{ij}}{\sum_{k \in \mathcal{S}_j \setminus \{i\}} g_{kj} \sum_{\ell \in \mathcal{R}_k \setminus \{j\}} x_{k\ell}(t)} \geq \gamma \qquad \text{if link } (i,j) \text{ is active at time } t, \text{ i.e., } x_{ij}(t) = 1. \tag{8.10}$$

The basic assumption of the model, i.e., the possibility of reducing the PER to a step function around the value $\gamma$, is indeed an approximation. However, it is much more accurate than those made under the protocol models. In fact, it takes into account physical propagation, and allows for a correct packet reception even in the presence of (moderate) interference, differently from the collision assumption. Also, it properly accounts for the cumulative character of interference. Indeed, the choice of $\gamma$ depends on the shape of the PER function, which in turn relates to the modulation scheme, and on the PER value which is considered as acceptable at the application level. However, none of these factors depends on MAC issues; thus, the physical model allows to operate between MAC and other layers in a more modular manner.

The drawback of this model is that it translates into more complex mathematical relationships than the protocol model. Moreover, if a specific MAC needs to be addressed, additional constraints are required. For example, in an IEEE 802.11 network, the physical model fails to describe certain constraints on link activation due to the RTS/CTS exchange, which are instead taken into account in the *11protocol* model. Hence, which is the best model to use ultimately depends on the purpose of the analysis. From the point of view of theoretical analysis of WMNs, however, the physical model has a good point against the protocol model, as described, e.g., in [10, 15]. In dense topologies, where the number of incoming or outgoing links at a node is high, the protocol models are very restrictive and obtain lower network parallelism, due to their requirement of silencing allegedly colliding connections. As shown above, when the WMN topology is rich of edges, all protocol models approach the *01protocol* model, which is the most restrictive case and implies that at most one edge is activated. This is a problem for WMNs, which, being meant

to provide good network coverage and high data rates, usually have a dense topology. The better performance obtained in this sense by utilizing the physical model should also imply the need to re-think existing access protocols for WMN. Indeed, the mesh versions of both IEEE 802.11 and IEEE 802.16 take these aspects into account. However, in our view the protocol design of improved interference-aware routing and scheduling strategies is still an open research challenge.

## 8.6 Performance Evaluation

In this section, we focus on the problem of defining efficient link activation patterns which not only satisfy all the constraints but also deliver traffic to the MAPs acting as gateways for the WMN. We focus on the minimal time scheduling problem, i.e., to deliver a given amount of traffic from all the non-gateway MRs to the MAPs (as we deal with the uplink case) in the shortest possible time. This problem is also closely related to the throughput maximization, i.e., to obtain the highest amount of traffic delivered to the gateways in an assigned time. Indeed, with minor modifications our framework can work to solve this problem as well.

In the following, we will refer to the backlog queue length at node $i$, assumed to be varying over time, as $q_i(t)$. Thus, all non-gateway MRs have, at time 0, a backlog of length $q_i(0)$ to be sent to any of the MAPs. The minimal time scheduling problem corresponds to finding the lowest length $T_{\min}$ of a feasible link activation pattern which delivers all traffic to the gateways. Denoting the set of gateways by $\mathcal{Y}$, this implies that

$$T_{\min} = \min\{t : q_i(t) = 0\,,\ \forall i \in \mathcal{N} \setminus \mathcal{Y}\}. \tag{8.11}$$

For simplicity, we assume that the value of $q_i(0)$ is known *a priori* and no further packet arrivals take place after link activation has started. In this way, if the uplink problem can be solved over a specified finite time-horizon $T$, i.e., $T_{\min}$ is lower than or equal to $T$, its solution can also serve as the basis for a periodic schedule, where a link activation pattern of length $T$ is indefinitely repeated. In other words, it is possible to see the uplink problem as a way to deliver a given amount of packets under loose delay guarantees (i.e., every packet is delivered within $2T$ slots, provided that the arrival rate to the MRs from MCs can be assumed constant). A further extension is possible to the cases of prioritized traffic with different priority classes or different required delay guarantees. Another option is to consider packet arrivals within the time frame. All these differences do not change most of the considerations we will present in the following, and can be investigated within a similar framework. We identify them as possible interesting directions for future research.

Finding the shortest-time link activation pattern for the uplink problem can be addressed in the context of an optimization problem, by adding proper *flow constraints* to the already mentioned duplex and interference constraints. Among these, the most important is the flow conservation property, i.e., the traffic transmitted over $(i, j)$ in a given time slot $t$ is upper bounded by the number of packets available at node $i$ after the transmission occurred at time $t - 1$. Note that this is true if, as assumed above, all traffic arrives at the MRs at the beginning of the schedule.

Other conditions may impose that an edge $(i, j)$ can be active at time $t$ only if $q_i(t) > 0$ or that the edges *from* an MAP are never activated. These constraints are not strictly necessary, but they eliminate from the feasible region parts of the search space which are guaranteed to contain only non-optimal solutions.

Several approaches have been proposed to formalize the problem of finding the optimal link activation pattern to minimize the time to deliver all traffic to MAPs [7, 26]. However, the resulting optimization problem is NP-complete [10]. For this reason, we focus our analysis on some theoretical results on the overall performance of WMNs for the minimal scheduling problem. Within this approach, not only is it possible to frame other existing results, but also we are able to draw interesting guidelines and conclusions about the performance of WMNs.

### 8.6.1 Theoretical Performance Bounds

Determining the value of $T_{\min}$ is interesting for both theoretical and practical reasons. In fact, the problem of delivering a given amount of traffic can also be seen from the information theoretical point of view as a capacity estimation, since the shorter the time to deliver a given amount of packets, the higher the throughput over a given time interval. Also, if $T_{\min}$ is sufficiently low, a centralized periodic scheduling can be implemented. However, the problem of determining $T_{\min}$ exactly is very complicated. Not only is it an NP-complete problem, but also it strongly depends on the network parameters, i.e., the graph topology, the edge rates and the initial backlog at each node.

Thus, solutions based on integer linear programming often introduce simplifications to make the problem more tractable. For instance, [13] employs a *fluidic approximation* to the link rates, i.e., the $x_{ij}$ variables are relaxed to be time-invariant real numbers between $0$ and $1$ instead of being binary digits variable over time. In other words, the $x_{ij}$ variables represent the average activity of link $(i, j)$ over the time period. However, this approach has some drawbacks, for example it leads to rounding problems. If it is found that the optimal average link activity for link $(i, j)$ is, say, $0.83$, and $T$ is found to be equal to $10$, it is not clear whether $(i, j)$ should be active on $8$ or $9$ time slots. Moreover, the practicality of the approach is decreased with respect to the initial integer problem, where the solution could be directly translated into a schedule simply by taking the resulting link activation pattern, which is no longer possible. Finally, the overall $T_{\min}$ to schedule all the traffic is underestimated with respect to the original integer case, as observed by the authors themselves.

Another possibility which is sometimes proposed [7] is to employ topology control to reduce the number of edges which can be activated. Even though this indeed decreases the complexity of the problem, we argue that this procedure can lead to a severe decrease of the transmission parallelism, therefore obtaining low throughput as a result. Thus, it is in general not recommended to prune edges to decrease the cardinality of $\mathcal{E}$. This is true even for the cases where topology control is claimed to be interference-aware: as discussed in previous sections, allocating non-interfering simultaneous connections is a task to be performed at the MAC layer, i.e., through a scheduler (or in our case, through a joint routing-scheduling procedure), not with a

routing algorithm. If interference awareness is introduced in the network by simply reducing the possible routes, the most significant result is a decrease of the overall performance.

Finally, the most natural way to deal with difficult problems, i.e., to introduce a heuristic solution method, is also common in the literature [6,27]. Indeed, to identify novel and possibly topology-adaptive heuristics or meta-heuristics is another possible direction for further research. Instead of proposing yet another heuristic, we present some theoretical results which hold in general for WMNs. Similar findings have been also presented in other contributions [13, 15, 17], which however heavily rely on the assumption of the protocol interference model. Instead, our analysis is *independent of the underlying interference model*, as it only relies on the half-duplex assumption, which, as discussed in Subsection 8.4.2, holds in any case. Under this hypothesis, we derive theoretical bounds for the performance of WMNs, in which the interference model of choice can be framed (obtaining different results, according to how restrictive it is).

Prior to describing the analytical formulation, note what follows about the notation of the following theoretical statements. As observed above, a TDMA scheduler operates on discrete time slots. Thus, the *number of slots* required to accomplish a transmission is an integer. However, in the following we will refer to the *time to transmit* a given amount of traffic as a real number. Different from [13], this does not imply that we are relaxing the constraints of $x_{ij}(t)$ to be integer, but simply that in all the cases where the time to transmit is non-integer, the number of slots corresponds to its rounded-up version. Thus, the results shown in the following can be refined to properly capture the fact that time slots are integer numbers by adding ceilings where necessary.

The first result is an upper bound on $T_{\min}$ which can be seen, to some extent, as introduced by [17]. The overall idea of this paper is to determine the minimal time scheduling by deriving maximal cliques of edges which can be compatibly allocated. This is just a different formulation of the problem, which does not solve in any way its NP-completeness. Besides, the whole analysis is based on the protocol interference model. However, an interesting point is given in the paper. No matter how inefficient the schedule is, at least one edge should be activated at a time. Thus, an upper bound for $T_{\min}$, denoted as $T^U_{\min}$, is implicitly obtained, exactly by taking this as a worst case assumption. Note that this upper bound corresponds to what, in Subsection 8.5.3, was referred to as *01protocol* model. Therefore, this upper bound is also tight in the sense that there is an interference condition in which the shortest link activation pattern must necessarily be $T^U_{\min}$ slots long. This is exactly when the interference corresponds to the *01protocol* model, which is the worst possible case.

To derive $T^U_{\min}$, we simply take a weighted shortest path to the gateways from any node, e.g., by using the well known Dijkstra algorithm, where the weights are the inverse of the link rates $r_{ij}$. In fact, it is easy to see that, if transmitting a backlog $q$ over a link of rate $r$ takes a time equal to $qr^{-1}$, the time to transmit it over the series of two links (activated one at a time) having rate $r_1$ and $r_2$, respectively, is $q(r_1^{-1} + r_2^{-1})$. Finally, the value of $T^U_{\min}$ is derived as

$$T^U_{\min} = \sum_{i \in \mathcal{N} \setminus \mathcal{Y}} q_i(0) \sum_{e \in \mathcal{P}_i} r_e^{-1} \tag{8.12}$$

where $\mathcal{P}_i$ is the shortest path in the sense mentioned above for node $i$.

The upper bound described by $T^U_{\min}$ corresponds to a very conservative case of protection against interference. The minimal time to deliver all the traffic is $T^U_{\min}$ only in the case of *01protocol* model, or very high SIR target in the physical model. For this reason, we introduce also a lower bound on $T_{\min}$ for a *single gateway case*. This condition is likely to be present in most WMNs and has been first envisioned by [15] as a possible bottleneck for the network capacity of such systems. The authors of this paper argue that if a single gateway is used and the highest rate of all links entering in it is $a$, there is a lower bound on the time to deliver all packets, equal to $Q/a$, where $Q$ is the sum of all backlogs in the network at time 0. This lower bound is trivial in most cases, but it might be interesting for certain sparse topologies. Especially in [15] it is shown that a chain topology behaves badly in this sense. Moreover, the authors further improve this result by giving some theoretical considerations based on the protocol model (more specifically, the *11protocol* model version).

Though inspired by this result, we follow here another approach. We demonstrate that taking only the half duplex constraint into account is sufficient to significantly improve the aforementioned lower bound. Even though the problem of determining a tight lower bound on $T_{\min}$ would still be NP-complete, we remark that in practice our theoretical result gives a good estimate of $T_{\min}$ for the case of no interference (hence, the other extreme with respect to $T^U_{\min}$) in several cases. If more than one gateway is present, the gateway bottleneck is strongly mitigated; thus, an immediate conclusion of our analysis is that WMNs perform significantly better if two or more MAPs are available.

To derive the lower bound, referred to in the following as $T^L_{\min}$, we proceed as follows. Consider, as in [15], the edge entering the gateway with highest rate (equal to $a$). The transmitter node of this edge would be called in the following "MR number 1" and its backlog will be denoted as $q$. As above, let $Q$ indicate the overall backlog in the network. Let $s$ be the highest rate of all edges entering MR number 1, and let $b$ be the highest rate among the edges entering the gateway, not counting the one from MR number 1 (hence, $a \geq b$). If multiple nodes can be chosen as MR number 1, since several edges to the gateway have equal rate, simply put $b = a$ and $s$ will be consequently equal to the highest possible rate among all edges entering those nodes. For simplicity, we assume that $b > 0$ and $s > 0$. However, it is still possible to generalize the result shown in the following to $b = 0$ or $s = 0$.

The situation is represented in Fig. 8.4. In the following, we neglect all edges in the rest of the network, and we will also neglect multiple edges with identical rates. As a matter of fact, we only consider three links: from MR number 1 to the gateway, from the rest of the network to the gateway and from the rest of the network to MR number 1, having rates $a$, $b$ and $s$, respectively. With a slight abuse of notation, we will call them with their rate value, for brevity. The lower bound $T^L_{\min}$ is derived considering all the traffic in the rest of the network (equal to $Q - q$ packets) to be always available on "border" nodes which can use these links. Actually, this is an

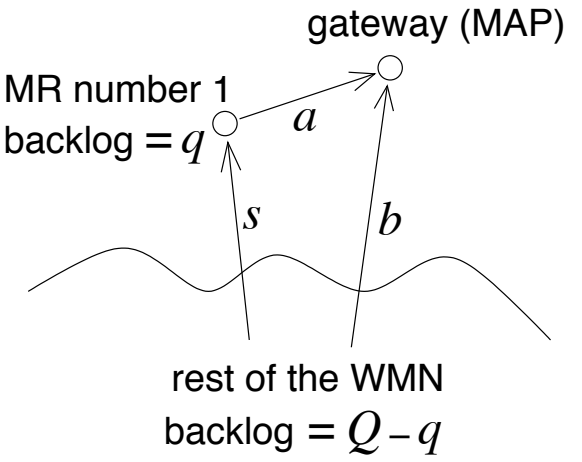



**Fig. 8.4.** Notations used to derive the lower bound $T^L_{\min}$.

optimistic assumption as these packets can be instead queued at other nodes which are not directly connected to the gateway, or to MR number 1. The derivation of $T^L_{\min}$ is obtained through the following theorem.

**Theorem 1.** *A lower bound on* $T_{\min}$ *is given by*

$$T^L_{\min} = \frac{q}{a} + \frac{Q-q}{s+b}\left(1+\frac{s}{a}\right). \tag{8.13}$$

*Proof.* First, observe a general property. When edge $a$ is active, i.e., MR number 1 sends packets to the gateway, no other transmission to these nodes can be activated due to the half-duplex constraint. At most, it is possible to activate in parallel some transmissions within the "rest of the network," but this has no effect whatsoever, since we are under the optimistic assumption that all the traffic which is not queued at MR number 1 is always available for transmission on links $b$ and $s$. We can therefore neglect these transmissions as they can not improve the lower bound $T^L_{\min}$. Thus, it is not restrictive to assume that all the traffic available at MR number 1 is transmitted first, which takes a time equal to $q/a$. Then, $T^L_{\min} = q/a + T_1$, where $T_1$ is a lower bound on the delivery time in the same network, where however $q$ has been delivered to the gateway. Since MR number 1 now has no packets in the queue, links $b$ and $s$ need to be activated. The best possibility (i.e., the one minimizing the delivery time) is that they can operate perfectly in parallel. During such a parallel transmission, assume that $x$ and $y$ are the amounts of traffic sent over link $s$ and $b$, respectively. After this transmission, no packets are left in the rest of the network, so $x+y = Q-q$. Moreover, the minimum transmission time is obtained when $x$ and $y$ take exactly the same time to be transmitted. This means that

$$\begin{cases} x + y = Q - q \\ x : s = y : b. \end{cases} \tag{8.14}$$

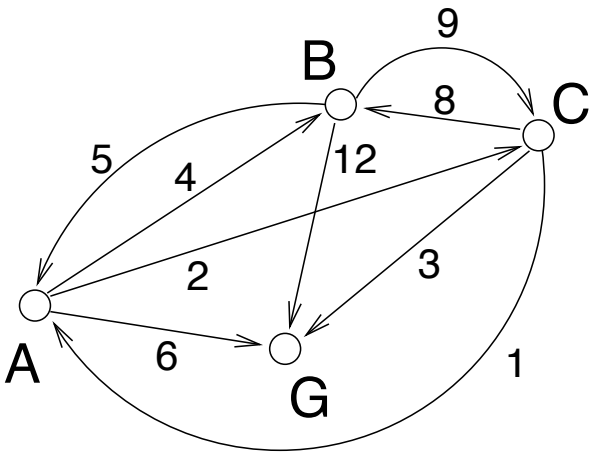



**Fig. 8.5.** An example of network topology.

This system of equations can be solved so as to obtain

$$x = s\frac{Q-q}{s+b}\,, \quad y = b\frac{Q-q}{s+b}\,. \tag{8.15}$$

The parallel transmission over $b$ and $s$ is also found to have a duration of $(Q-q)/(s+b)$. After its termination, an amount of traffic equal to $y$ has been delivered to the gateway, whereas $x$ is still in queue at MR number 1. The best possibility to transmit $x$ is to use link $a$, which takes a time equal to $(s/a)(Q-q)/(s+b)$. Thus, collecting all these results,

$$T_1 = \frac{Q-q}{s+b}\left(1+\frac{s}{a}\right) \tag{8.16}$$

and the theorem is proved. ■

This result has many practical consequences. For example, not only does $a$ impact $T_{\min}$, but so do $b$ and $s$. In particular, if $s \ll a$, the gateway bottleneck is worsened, since packets arrive at MR number 1 with very low rate, thus alternate paths (hence with rate $b$ lower than $a$) have to be used. On the same line, link $a$ can not always be used for transmitting packets, since it can not be activated when MR number 1 is receiving. If $b$ is considerably lower than $a$, there may be a decrease in the network throughput. These considerations give some practical guidelines for network deployment. First of all, it is important to have several "good" links to the gateway, i.e., $b$ should be close to $a$, and there should be multiple non-interfering paths to the gateway, so as to allow parallel allocation of links to the gateway and to some of the neighbors of the gateway. Instead, if all routes to the gateway traverse the

same node, the single gateway bottleneck is worsened. Moreover, the rate of connections to the gateway should be high, but it is also important to have a good relaying speed to the gateway neighbors (i.e., high $s$).

For what concerns numerical evaluations, we found that this lower bound works well in practical cases. In particular, it is much stricter than the trivial lower bound given by $Q/a$, and it also has the advantage of limiting the analysis to three numerical values, i.e., the best rate and the second best rate of edges entering the gateway and the best rate of edges entering MR number 1. To give an idea of this, consider[3] the sample WMN represented in Fig. 8.5. Numbers reported on the edges denote their rates (again, with the same abuse of notation, we speak, e.g., of edge 1 to indicate the one from node C to node A). Assume that node G is the gateway (this is also implicitly taken into account in the figure, where no edges from G are depicted), and that $q_i(0) = 24$ for every node. In such a case, $T^U_{\min} = 9$ (shortest paths are through direct links for all MRs but for node C, whose best path is through edges 8 and 12, with an overall rate of $24/5$), which is indeed the actual value of $T_{\min}$ if the *01protocol* model holds. Since $a = 12$, $b = 6$, $s = 8$, the "trivial" lower bound $Q/a$ is 6, whereas $T^L_{\min} = 7.71$. This latter value is much more accurate than the former, as the minimal length of the schedule is 8 time slots (the optimal schedule corresponds in fact to activating edges 8 and 6 simultaneously for three slots, then edge 6 alone for one slot and finally edge 12 for four slots). The slightly lower value of $T^L_{\min}$ with respect to the real value is a consequence of the fact that the parallelism of edges 8 and 6 is not perfect (even though the round-up still eliminates this issue). Moreover, the activation of this parallel transmission is possible only if it does not violate any additional interference constraint. In fact, if for example the *11protocol* interference model is assumed, since the network topology here is a clique, we obtain $\mathcal{I}(e) = \mathcal{E}$ for all $e$, thus we fall again in the case described by the upper bound.

This suggests that, according to the interference model, the network performance in terms of $T_{\min}$ moves from the upper to the lower bound (or close to it). In the next subsection, we will show a more extensive analysis of this behavior.

### 8.6.2 Numerical Results

In this section we analyze, through an example, the resulting network capacity when different interference models are adopted. To this end, we focus on the network topology reported in Fig. 8.6 which consists of six nodes. Node 0 is assumed to be the gateway. The links between nodes are represented by directional edges whose rate is either $\alpha$ or $\beta$. With respect to protocol models, the disturbance range is assumed to be equal to the transmission range, i.e., $\vartheta = 1$. Thus, a node can disturb only the nodes it can transmit to, and vice versa. The nodes that do not have a direct link toward the gateway exploit their neighboring nodes to relay their packets. We assume that each node is provided with an equal amount of traffic to forward to the

[3] We remark that this topology, and also the one shown in the next subsection, have only the value of examples and are not proposed in this paper as realistic or efficient network deployments.

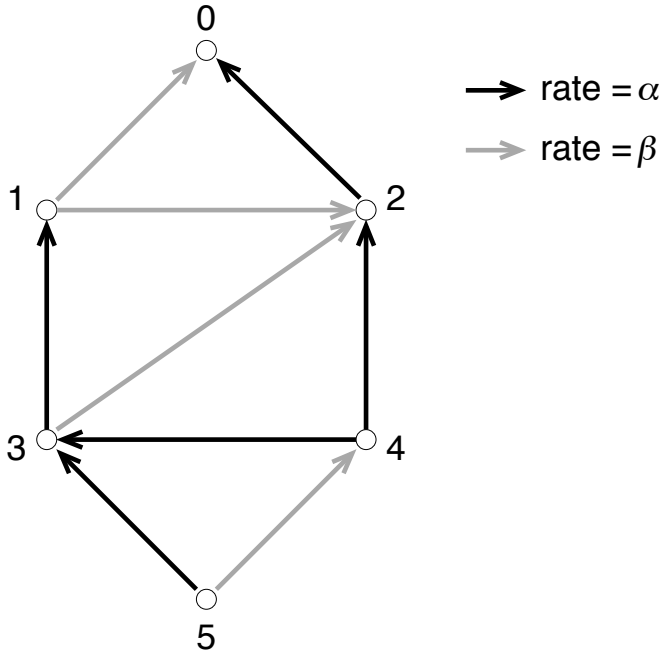



**Fig. 8.6.** Example topology: The gateway is node 0, edge rates are either $\alpha$ (black links) or $\beta$ (grey links).

gateway. The network capacity is evaluated by means of $\lceil T_{\min}/q \rceil$, i.e., the number of slots needed to deliver the overall network workload to the gateway, normalized over the initial amount of traffic of the nodes. To some extent, we can draw in this way general conclusions on the network capacity irrespective of the initial traffic load of the nodes. Depending on the interference model, our analysis is carried out either with the theoretical results described in Subsection 8.6.1 or through numerical evaluations, performed with an exhaustive search over all possible link activation patterns.

In Fig. 8.7 we plot this metric versus the ratio $\alpha/\beta$, considering a case where $\alpha\beta = 1$, for different interference models. As can be seen, the *01protocol* model curve (also corresponding to the analytical upper bound for the performance of any MAC) always lies above the other ones due to the restrictive constraint that at most one link can be active at a time. Even though the curves exhibit slightly variable behavior when $\alpha/\beta$ is changed, there is in any case a significant gap (a factor of $1.8$ or more) between the *01protocol* model and the theoretical lower bound curves. Instead, the half duplex performance, derived through exhaustive search in the least restrictive condition of simultaneous link activation, is well approximated by the analytical lower bound. Especially, the lower bound is fairly tight when $\alpha \leq \beta$.

Any value between the *01protocol* model and the half-duplex curves may potentially be achievable depending on the interference model. In particular, if the physical

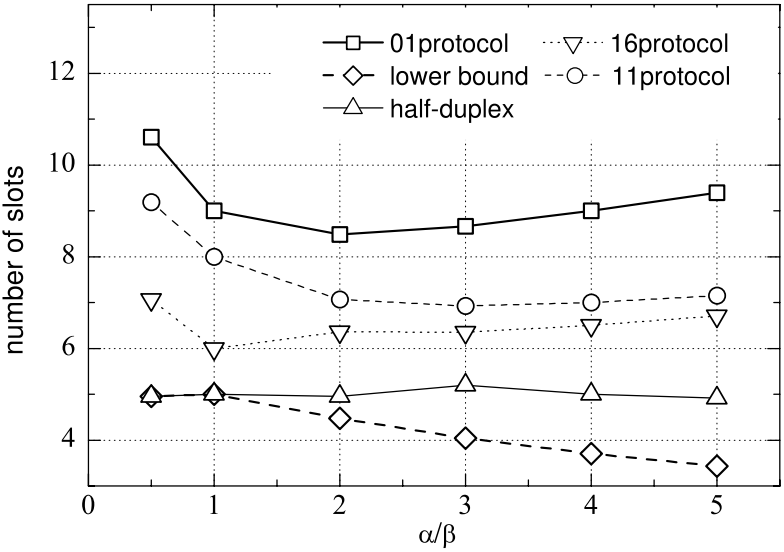



**Fig. 8.7.** Numerical results (for all protocol models $\vartheta = 1$ is considered).

model is utilized, the performance will span between the two extreme curves almost with continuity. If a protocol model is used instead, the behavior is more difficult to change. The performance of both *11protocol* and *16protocol* can be observed in the figure to be according to the reasonings of Subsection 8.5.3. For example, the *11protocol* is closer to the *01protocol* than the *16protocol* due to its more restrictive assumptions. Recall that, in the *11protocol*, all the nodes falling into the disturbance range of both the transmitter and the receiver need to be silenced. This assumption is relaxed in the *16protocol* which thus permits more parallelism of the transmissions. However, both the *11protocol* and *16protocol* models obtain a significantly higher $T_{\min}$ than when only the half-duplex constraint is assumed, but no further constraint is imposed.

The condition $\alpha = \beta$ corresponds to the case where the link rates in the network become homogeneous, i.e., all nodes transmit at the same rate. As envisioned in the previous discussion, this is a good condition for achieving high throughput, since the lack of potential bottlenecks caused by slower links permits to efficiently exploit the overall network capacity. As soon as the link rates in the network become heterogeneous, we observe that the performance degrades in this sense. Note that, because $\alpha\beta = 1$, when $\alpha$ is increased, it also happens that $\beta$ decreases in an inversely proportional fashion. For $\alpha > \beta$ link rates are higher on aggregate: e.g., when $\alpha/\beta = 4$, the average link rate is equal to $4/3$ instead of 1. However, this does not correspond to an improvement in the scheduling efficiency, because there are high rate links to the gateway, but also strong variability of the link rates is present. Due to the aforementioned bottlenecks, the case only constrained by half-duplex limitations keeps its performance almost constant.

## Conclusion

In this chapter, we have investigated some research issues arising in the context of link activation for WMNs. Specifically, we have revisited the classic problems of routing and link scheduling over multi-hop wireless networks to provide the reader with a clear overview of the hottest topics in WMNs. After a brief introduction discussing preliminary concepts of wireless networks, we have critically reviewed the recent literature in this field, highlighting pros and cons of possible approaches to the problem. We have proposed an approach which jointly considers the routing and link scheduling problems. To this aim, we have introduced and described theoretical models to characterize wireless networks, which include the nodes' transmission/reception constraints and the interference of wireless links. Within this theoretical framework, we have discussed the characteristics of the most common MAC protocols. Finally, we have derived theoretical performance bounds for network capacity and have compared these bounds to the results obtained by the presented models in a sample topology.

We believe that these results can be useful in many ways. Certainly, one possibility is to use them as guidelines for WMN deployment so as to avoid bottlenecks in the network and allow instead high data rates to the end users, which corresponds to the major objective of such systems. At the same time, we remark how our findings highlight the need for a proper interference characterization. This point seems often neglected, as testified by the widespread usage of the protocol interference models, which do not correctly describe the underlying physical aspects and also lead to pessimistic performance results. We believe that further investigation on wireless interference models and their impact on MAC and network layer aspects is an interesting scientific challenge. Finally, efficient link activation strategies exhibit a performance with a high degree of variability between the evaluated analytical upper and lower bounds. In this respect, novel proposals which are able to fill this gap are clearly emphasized as very promising directions of future research.

## References

1. I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *Computer Networks (Elsevier)*, vol. 47, pp. 445–487, Mar. 2005.
2. R. Bruno, M. Conti, and E. Gregori, "Mesh networks: Commodity multihop ad hoc networks," *IEEE Commun. Mag.*, vol. 43, pp. 123–131, Mar. 2005.
3. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.
4. *Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16, 2004.
5. K. N. Ramachandran, M. M. Buddhikot, G. Chandranmenon, S. Miller, E. M. Belding-Royer, and K. C. Almeroth, "On the design and implementation of infrastructure mesh networks," in *Proc. WiMesh*, Santa Clara, CA, USA, Sept. 2005.
6. M. Alicherry, R. Bhatia, and L. B. Li, "Joint channel assignment and routing for throughput optimization in multiradio wireless mesh networks," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1960–1971, Nov. 2006.

7. J. Tang, G. Xue, and W. Zhang, "Interference-aware topology control and QoS routing in multi-channel wireless mesh networks," in *Proc. ACM MobiHoc*, Urbana-Champaign, IL, USA, May 2005, pp. 68–77.
8. R. L. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multihop wireless networks," in *Proc. IEEE INFOCOM*, vol. 1, San Francisco, CA, USA, 2003, pp. 702–711.
9. H. Balakrishnan, C. L. Barrett, V. S. A. Kumar, M. V. Marathe, and S. Thite, "The distance-2 matching problem and its relationship to the MAC-Layer capacity of ad hoc wireless networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1069–1079, Aug. 2004.
10. G. Brar, D. Blough, and P. Santi, "Computationally efficient scheduling with the physical interference model for throughput improvement in wireless mesh networks," in *Proc. ACM MobiCom*, Los Angeles, CA, USA, Sept. 2006, pp. 2–13.
11. L. Tassiulas and A. Ephremides, "Jointly optimal routing and scheduling in packet ratio networks," vol. 38, no. 1, pp. 165–168, Jan. 1992.
12. M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless mesh networks with orthogonal channels," *IEEE/ACM Trans. Networking*, vol. 13, no. 4, pp. 868–880, Aug. 2005.
13. ——, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM MobiCom*, Cologne, Germany, 2005, pp. 73–87.
14. K. Xu, M. Gerla, and S. Bae, "How effective is the IEEE 802.11 RTS/CTS handshake in ad hoc networks," in *Proc. IEEE Globecom*, Taipei, Taiwan, R.O.C., 2002, pp. 72–76.
15. J. Jun and M. L. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Commun. Mag.*, vol. 10, no. 5, pp. 8–14, Oct. 2003.
16. W. Wang and X. Liu, "A framework for maximum capacity in multi-channel multi-radio wireless networks," in *Proc. IEEE CCNC*, vol. 2, 2006, pp. 720–724.
17. N. Ben Salem and J.-P. Hubaux, "A fair scheduling for wireless mesh networks," in *Proc. WiMesh*, Santa Clara, CA, USA, 2005, invited paper.
18. P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.
19. C. Reis, R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Measurement-based models of delivery and interference in static wireless networks," in *Proc. ACM SIGCOMM*, Pisa, Italy, 2006, pp. 51–62.
20. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. ACM MobiCom*, Philadelphia, PA, USA, July 2004, pp. 114–128.
21. Y. Yang, J. Wang, and R. Kravets, "Designing routing metrics for mesh networks," in *Proc. WiMesh*, Santa Clara, CA, USA, Sept. 2005.
22. H. Wu, F. Yang, K. Tan, J. Chen, Q. Zhang, and Z. Zhang, "Distributed channel assignment and routing in multiradio multichannel multihop wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1972–1983, Nov. 2006.
23. T. Salonidis and L. Tassiulas, "Distributed on-line schedule adaptation for balanced slot allocation in wireless ad hoc networks," in *Proc. IEEE IWQoS*, Montreal, QC, Canada, June 2004, pp. 20–29.
24. P. Djukic and S. Valaee, "Distributed link scheduling for TDMA mesh networks," in *Proc. IEEE ICC 2007*, Glasgow, Scotland.
25. K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Networks*, vol. 11, no. 4, pp. 471–487, July 2005.

26. M. Cao, W. Ma, Q. Zhang, X. Wang, and W. Zhu, "Modelling and performance analysis of the distributed scheduler in IEEE 802.16 mesh mode," in *Proc. ACM MobiHoc*, Urbana-Champaign, IL, USA, May 2005, pp. 78–89.
27. H. Wei, S. Ganguly, R. Izmailov, and Z. J. Haas, "Interference-aware IEEE 802.16 WiMax mesh networks," in *Proc. IEEE VTC 2005 Spring*, Stockholm, Sweden.
28. A. P. Subramanian, M. M. Buddhikot, and S. Miller, "Interference aware routing in multi-radio wireless mesh networks," in *Proc. WiMesh*, Reston, VA, USA, 2006, pp. 55–63.
29. S. O. Krumke, M. V. Marathe, and S. Ravi, "Models and approximation algorithms for channel assignment in radio networks," *Wireless Networks*, vol. 7, no. 6, pp. 575–584, Nov. 2001.
30. P. Kyasanur and N. H. Vaidya, "Capacity of multi-channel wireless networks: Impact of number of channels and interfaces," in *Proc. ACM MobiCom*, Cologne, Germany, 2005, pp. 43–57.
31. A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "A multi-radio unification protocol for IEEE 802.11 wireless networks," in *Proc. IEEE BroadNets*, San José, CA, USA, 2004, pp. 344–354.
32. A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh, "Partially overlapped channels not considered harmful," in *Proc. ACM SIGMetrics/Performance*, Saint Malo, France, 2006, pp. 63–74.
33. F. M. E. Amaldi and A. Capone, "Planning UMTS base station location: Optimization models with power control and algorithms," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 939–952, Sept. 2003.
34. B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, "Gateway placement optimization in wireless mesh networks with QoS constraints," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2127–2136, Nov. 2006.
35. A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Addison Wesley Longman Publishing Co., Inc.
36. J. W. Wallace and M. A. Jensen, "Modeling the indoor MIMO wireless channel," *IEEE Trans. Antennas Propagat.*, vol. 50, no. 5, pp. 591–599, May 2002.
37. A. Scaglione and Y.-W. Hong, "Opportunistic large arrays: Cooperative transmission in wireless multihop ad hoc networks to reach far distances," *IEEE Trans. Signal Processing*, vol. 51, no. 8, pp. 2082–2092, Aug. 2003.
38. Y. Sagduyu and A. Ephremides, "Crosslayer design for distributed MAC and network coding in wireless ad hoc networks," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, 2005, pp. 1863–1867.
39. L. Lai, K. Liu, and H. E. Gamal, "The three-node wireless network: Achievable rates and cooperation strategies," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 805–828, 2006.
40. S. M. Das, H. Pucha, D. Koutsonikolas, Y. C. Hu, and D. Peroulis, "DMesh: Incorporating practical directional antennas in multichannel wireless mesh networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2028–2039, Nov. 2006.
41. J. A. Stine, "Exploiting smart antennas in wireless mesh networks using contention access," *IEEE Wireless Commun. Mag.*, vol. 13, no. 2, pp. 38–49, Apr. 2006.
42. D. Kotz, C. Newport, R. S. Gray, J. Liu, Y. Yuan, and C. Elliott, "Experimental evaluation of wireless simulation assumptions," in *Proc. ACM MSWiM*, Venice, Italy, 2004, pp. 78–82.
43. S. Basagni, M. Conti, S. Giordano, and I. Stojmenović, Eds., *Mobile Ad Hoc Networking*. New York: IEEE press and John Wiley & Sons, 2004.
44. H. Zhai and Y. Fang, "Medium access control protocols in mobile ad hoc networks: Problems and solutions," in *Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless*

*and Peer-to-Peer Networks*, J. Wu, Ed. Boca Raton, FL, USA: Auerbach Publications, Taylor & Francis Group, 2006, ch. 15, pp. 231–250.
45. A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?" in *Proc. QShine*, no. 2, Waterloo, ON, Canada, 2006, invited paper.
46. G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 535–547, 2000.

# 9

# Quality-Aware Routing Metrics in Wireless Mesh Networks

C. E. Koksal

The Ohio State University, USA
koksal.2@osu.edu

## 9.1 Introduction

In this chapter we address the problem of selecting good paths in networks made up of multiple wireless links[1], such as wireless mesh networks. By "good paths", we mean paths that both benefit individual data transfers (in terms of TCP connection throughput, for example), and which lead to high aggregate network capacity.

Finding good paths between nodes in a wireless network involves two steps:

1. *Assigning cost metrics to links and paths.*
2. *Disseminating routing information.*

The second step, route dissemination, has received much attention over the past decade. The link and/or path metrics need to be disseminated to the nodes in the network using a routing protocol, to help nodes select best paths in a distributed fashion. There are two types of protocols in how the route dissemination is done: proactive and reactive protocols.

Proactive protocols determine paths before there is any demand for communication. They calculate the routing tables ahead of time and maintain them through periodic update messages. Examples include Destination-Sequenced Distance Vector Routing (DSDV, [1]), Fisheye State Routing (FSR, [2]), and Optimized Link State Routing (OLSR, [3]).

Reactive protocols, on the other hand, do not calculate routes ahead of time. Route discovery follows the communication request. Examples of reactive protocols include Ad Hoc On Demand Distance Vector (AODV, [4]), Temporarily Ordered Routing Algorithm (TORA, [5]) and Dynamic Source Routing (DSR, [6]).

In this chapter we address the first issue, assigning cost metrics to links. Regardless of whether a protocol is proactive or reactive, it requires a mechanism to differentiate between different paths. This differentiation is done using cost metrics.

The cost metric of a link is the cost of forwarding a packet along the link. The problem of defining a cost metric is considerably harder in wireless networks than

[1] We use the term "link" to refer to the communication channel between a pair of nodes.

in traditional wired networks, because the notion of a "link" between nodes is not well-defined. The properties of the radio channel between any pair of nodes vary with time, and the reliable radio communication range is often unpredictable. The communication quality of a radio channel depends on background noise, obstacles and channel fading, as well as on other transmissions occurring simultaneously in the network. The appropriate cost metric must take into account a number of factors due to the vagaries of radio channels, which in turn makes the task of assigning metrics non-trivial. Moreover, it is desirable that the metrics for the links along a path be *composable*, so that the end-to-end cost of a path can be easily derived from the individual metrics of the links along the path.

We observe that the type of quality aware routing metric to be chosen depends on the physical layer being used. Designing and implementing a physical layer that can fully "hide" the vagaries of the radio channel from higher layers has proven to be difficult for a number of reasons. It requires the physical layer to be able to accurately estimate and adapt several parameters (e.g., transmit power, modulation, error control coding, etc.) to cope with channel conditions that vary rapidly in time. In fact, we are not aware of any current or next-generation radios that propose to employ sophisticated techniques to fully handle channel quality issues at the physical layer, because of implementation complexity and the absence of practically useful codes that can perform well (especially in the non-asymptotic limit of finite packet sizes) across the large range of channel conditions that are observed in practice.

Indeed, practical wireless radios such as the ones based on the various IEEE 802 standards (e.g., 802.11, 802.15, etc.) employ only a simple coding strategy, mostly for error detection. Nodes transmit at one of a discrete set of power levels, and rely on a small number of link-layer packet retransmissions to overcome errors. All other packet losses are visible to higher layers, where they may be recovered using end-to-end mechanisms (such as TCP retransmissions or packet-level forward error correction implemented by applications). Most wireless mesh networks are radio networks comprised of radios similar to 802.11.

Another way modern radios (e.g., 802.11 chip-sets) cope with channel variations is the use of adaptive modulation schemes, allowing higher layers to set one of several possible bit rates. If frame loss rates at a particular bit rate rise, reducing the bit rate can reduce the observed frame loss ratio and improve throughput. Several bit rate adaptation schemes have been proposed (see [7] for a detailed treatment), and the topic remains an active area of work. We view bit rate selection as being complementary to quality-aware routing, in the sense that once the routing protocol picks the best neighbor to use for a destination using measured cost metrics, the link layer picks the best bit rate (modulation scheme) to use for that neighbor.

In the presence of bit rate adaptation, some routing metrics may need to be readjusted, and properly normalized with respect to the transmission rate. For instance, for many applications, a packet loss rate of 10% at 10 Mbps may be preferable over a 5% loss rate at 1 Mbps. Hence, a metric based solely on the loss rate should be modified to take the variety of available rates into account.

## 9.2 Routing Metrics for Wireless Mesh Networks

In this section we study seven cost metrics, discuss their relative benefits and shortcomings, and whether they would be appropriate for wireless mesh networks. These metrics are Hop Count, Per-hop Round Trip Time (RTT, [8]), Per-hop Packet Pair Delay (PktPair), quantized loss rate [9], Expected Transmission Count (ETX, [10]), modified ETX (mETX, [11]) and Effective Number of Transmissions (ENT, [11]).

### 9.2.1 Hop Count

The traditional approach to routing in ad hoc wireless networks is minimum-hop (shortest-path) routing (e.g., [1, 5]). The hop count is the simplest cost metric, and the simplicity of it may be attractive for networks for which mobility is high. Indeed, all other cost metrics require a link-level quantity to be measured or estimated and this process takes time, during which the same quantity may alter significantly. Consequently, in mobile networks one may be forced to use the simple hop count, which requires minimal amount of measurement.

Although simple, minimum-hop routing inherently "quantizes" the state of a link into one of the two states, "up" or "down." In reality, the state of a wireless link is not in any one of the two states at any point in time. For instance, Fig. 9.1 illustrates the packet delivery ratio taken from a certain link in the Roofnet wireless mesh network [12]. Each node in the network has an 802.11b wireless card and an antenna. The transmission rate is set to a constant 11 Mbps. The delivery ratio was obtained by sending a sequence of 1500-byte broadcast packets, with the receiver keeping track of which packets were received successfully. The successful receipt or loss of a packet defines a binary random variable; each sample delivery ratio in the graph is the average of a window of 40 successive binary random variables. The window advances by 1 for each reported sample. Clearly, the loss rate is almost never 0 or 1, but most of the time it is in the "grey" area in between these two extremes.

It has also been illustrated in different platforms (e.g., [13]) that in the presence of link variability, which is a common phenomenon in wireless mesh networks, minimum hop fails to have a satisfactory performance.

### 9.2.2 Per-hop Round Trip Time

A delay-based link cost metric was proposed in [8]. This metric uses the measured average round trip time (RTT) seen by unicast probes between neighboring nodes. It is originally built as a part of a Multi-Radio Unification Protocol (MUP) -a channel assignment protocol for community networks. Its application as a routing cost metric was implemented in [13].

To measure the channel, a probe packet is broadcast every 500 ms. Upon receiving a probe packet, each neighbor responds immediately, but in a non-preemptive

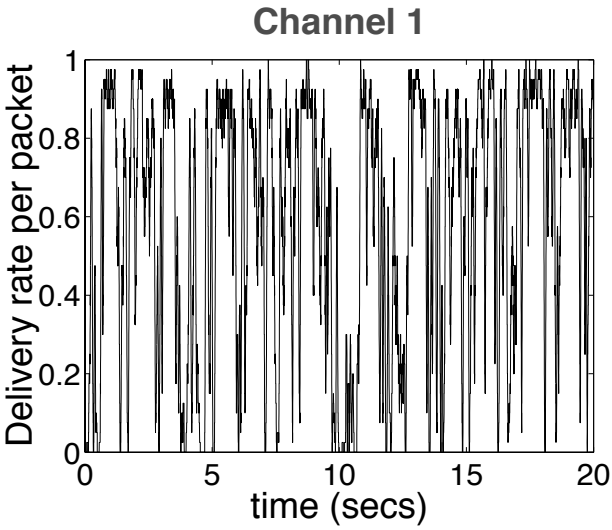



**Fig. 9.1.** Packet delivery rate for a link of Roofnet.

manner. The acknowledgment contains a time-stamp so that the RTT can be calculated. The node keeps an exponentially weighted moving average (EWMA) of the RTT samples for each neighbor:

$$\text{RTT estimate}[n+1] = 0.1 \times \text{RTT}[n] + 0.9 \times \text{RTT estimate}[n]$$

which is a low pass filter with a bandwidth of a few packets. The RTT estimate of the link is then assigned as the cost for the link. This metric is composable, since the sum of the RTT estimates over two links in cascade is the RTT estimate for the two-hop path.

The RTT cost metric contains several components contributing to the delay at a link.

- *Queueing delay*: Since the neighbors reply to the probe packets in a non-preemptive manner, the instantaneous RTT incorporates the time it takes for the existing jobs to be processed at a node.
- *Channel quality*: A packet may not be correctly decoded due to channel issues caused by fading or interference by other nodes not directly contending with our node. In this case, the packet is retransmitted up to a certain maximum number of times, contributing to the RTT calculation.
- *Channel contention*: If there are other nodes in the vicinity of one of the neighbors, the probe packet or the acknowledgment can get delayed due to direct contention. Contention can also be viewed as a channel issue (an outage) caused by a nearby node causing an intolerable amount of interference.

All of the above factors are legitimate factors that should be taken into account when considering the cost of a link. Indeed, it was illustrated in [8] that in a 12 node net-

work simulation with a real world web traffic model, the RTT metric is a reasonably well representative of the actual load at the nodes. Another set of simulations were run for a relatively lightly loaded network of 35 nodes, a small subset of which generates web traffic. When the RTT metric was used for channel assignment to pick the cleaner frequency for each hop, the network throughput increased by up to 70% and the average delay reduced by 50%.

However, there is a fundamental problem associated with using a routing metric, such as the RTT, which varies with varying load. It leads to either a highly oscillatory behavior or even instability. Specifically, suppose the delay at a certain node decreases due to reduced load at that node. Then, more and more of the paths tend to pass through this node, which will pull the delay, and hence the RTT metric back to a high value. The way the protocol is designed, such oscillations leading to route instability cannot be suppressed. The factors causing this type of route instability is referred to as "self interference."

In [13], the RTT metric was experimentally analyzed in a 23 node network in which every node pair initiates a long TCP session. The median of the average throughputs of all the sessions may be 75% lower when RTT is used instead of the simple hop count (which achieves around 1100 Kbps). The authors also illustrated that this reduction was indeed due to self interference, since the optimal path assignments change about 20 times more frequently with RTT, compared to the hop count.

One needs to be careful in using delay related quantities as a cost metric because of the self interference phenomenon. One solution proposed is to use another metric, per hop packet pair delay (PktPair), which is based on a simple modification to the per-hop RTT metric. We study the PktPair metric in the next section. Some other issues associated with the RTT metric can be listed as:

- The overhead associated with measuring the RTT may be high.
- This metric implicitly accounts for the link rate (the transmission time is inversely proportional to the link rate), but when the queueing delay is large relative to the transmission time, the link rate becomes an insignificant portion of the metric. However, in a dense network, increased link rate is a much more important component of the system performance since the interference and duration of contention are reduced by an increased data rate. Hence the amount of RTT spent on air should be a more important portion of the link cost compared to that spent in a queue at a node. Any throughput based metric can be modified simply to take the link rate into consideration, but it is not as easy for a delay based metric.
- This metric does not respond to the channel variability at time scales shorter than tens of packets. Indeed, the instantaneous RTT is sampled once every 500 msec and the resulting sequence is further low pass filtered with an EMWA filter. Thus, for a certain change to be effective in the route calculation, it should be sustained for an extended amount of time (5-6 seconds). The system is not responsive to the variations or bursty losses at time scales lower than that.

### 9.2.3 Per-hop Packet Pair Delay (PktPair)

PktPair was built by [13] in an effort to modify per hop RTT, which was shown to be problematic due to two issues. First one is the self interference and the second one is the relative significance of the queueing delay compared to the transmission time in the overall cost.

The idea of PktPair is based on sending a short probe packet ahead of a long one and using the short one to set a time reference. A small packet (of size 137 bytes) and a large one (1000 bytes) are sent in succession and each neighboring node keeps the time difference between the reception of these two packets. This value is fed back to the sender, which keeps an EWMA. This average is assigned as the cost metric for the link.

The measured difference between the times of reception of two successive packets includes potential delays due to contention for the medium with other nodes and the possible retransmissions due to channel issues caused by fading and other nodes communicating in the vicinity. Unlike the per hop RTT, PktPair does not have any component for the queueing and processing delay in it. This suppresses the route instability due to self interference to some extent. Indeed, the queueing and processing portion of an increase in delay do not contribute to an increase in the metric. However, an increase in contention still causes the metric to increase. Consequently, in a dense network with long term TCP flows, the average throughput increases (to 600 Kbps) by more than 100% and frequency of the change in the optimal path assignments reduce by about 50% compared to the RTT. Nevertheless the improvement is still not good enough for PktPair to outperform even the simple hop count metric.

Another issue associated with PktPair is the overhead, which is even higher than the overhead with per hop RTT.

### 9.2.4 Quantized Loss Rate

In [9] Yarvis *et al.* proposed a routing metric that estimates the per-link frame delivery ratios and uses the end-to-end path loss probability as the cost of routing over a path. Since the increase in the load affects the metric only through the increased contention, the effects leading to self interference are suppressed as much as the PktPair. The implementation was done for the sensor network platform; therefore, a large number of simplifications were made to make it practical in the presence of limited computational power.

To measure the link quality, each node keeps track of the number of correctly received packets from each of its neighbors. In particular, a window of the most recent 32 packets is considered for each downlink and an average number of correctly decoded packets is calculated. This value is then quantized depending on the region it lies: $Q_0$: 53-100% loss, $Q_1$: 21-53% loss, $Q_2$: 10-21% loss and $Q_3$: 0-10% loss. The midpoint of each region (i.e., 75%, 35%, 15% and 5%) is assigned as the representative of the region. Each node keeps track of its uplink to every neighbor as well and records the higher one of the two quantized loss rates as the (bi-directional) cost of the link.

The quantized loss rate metric is composable. Even though the end to end loss rate is not equal to the sum of the individual loss rates, one can simply use the $\log$ function. Indeed, we can add the -$\log$ of the estimated delivery rate ($R_e = 1 -$ loss rate) of each link to get the $\log$ of the end-to-end delivery rate for the path. This simple modification is used in the actual algorithm. The following table summarizes the metric assignment process.

| Quality | delivery rate | $R_e$ | $-\log(R_e)$ | cost metric |
|---|---|---|---|---|
| $Q_3$ | 90-100% | 0.95 | 0.05 | 1 |
| $Q_2$ | 79-90% | 0.85 | 0.16 | 3 |
| $Q_1$ | 47-79% | 0.65 | 0.43 | 8 |
| $Q_0$ | 0-47% | 0.25 | 1.39 | 28 |

This metric was tested over DSDV in a sensor network platform, and its performance is compared with that of the plain DSDV, for which the hop count is the cost metric. For 28 nodes, the quantized loss rate metric reduced the network wide loss rate by a percentage between 24-32%. For increased number of nodes, the amount of improvement decreases (e.g., for a 48 node network, percent improvement is between 6-20% and for a 91 node network it is between 2-4%). The authors argued that a good portion of this reduction in improvement might be due to the limitation of computational resources in the sensor nodes. Specifically, an increased number of nodes may be leading to an overflow in the neighbor lists, causing them to become ineffective. Note that, in wireless mesh networks, the lack of resources is less of an issue and the reduced improvement may be less significant.

Another issue about this metric is that it does not account for the total bandwidth consumed, because it will prefer two links of low loss rate over a single link with higher loss-rate. When link-layer retransmissions are used, the one-hop path may be able to deliver the packet without as many total transmissions as the two-hop path. In fact, ETX is motivated by this observation.

### 9.2.5 Expected Transmission Count (ETX)

ETX is a metric proposed by [10] for 802.11-based radios employing link-layer retransmissions to recover from frame losses. Basically, the ETX of a radio link is the estimated average number of $\{$data frame, ACK frame$\}$ transmissions necessary to transfer a packet successfully over the wireless link. In ETX, each node estimates the frame loss ratio $p_f$ to each of its neighbors over a recent time window, and obtains an estimate $p_r$ of the reverse direction from its neighbor. These loss estimates are obtained using broadcast probe packets (that are not retransmitted) at the link layer once every second. The estimate for $p_f$ and $p_r$ is, respectively, the fraction of the probes and the acknowledgments correctly decoded in the last ten seconds. The node then calculates the *expected transmission (ETX) count* for the link between the neighbor as $\frac{1}{(1-p_f)(1-p_r)}$. The ETX metric is composable, since the expected value of total number of transmissions over a path is the sum of the individual expected

number of transmissions of the links along the path. In the presence of bit rate adaptation, the only modification required for ETX is to use the Expected Transmission *Time* (rather than Count) as the metric [14], because a lower bit rate ends up using the channel for a longer period of time.

The number of transmissions of a packet on a radio link is an appealing cost metric because minimizing the total number of transmissions maximizes the overall throughput. It was shown in [13] that the ETX metric improves the average throughput of the TCP flows in the 23 node network (to 1357 Kbps) by 23.1% over the hop count metric. Also, the frequency of the changes in the calculated optimal paths is only 3 times as much as the hop count, which implies that the effects leading to self interference are mostly suppressed. This is expected since the link level retransmissions depend only on the link level packet errors caused by channel issues. The channel issues are almost completely independent of the load at a node.

Although the experimental results show that ETX performs better that traditional shortest path routing under static network conditions, it may perform poorly under highly variable channel conditions and burst-loss situations. Indeed, the ETX of the link is the reciprocal of the (estimated) probability of correct packet delivery. This definition implies that the probability of delivery of distinct packets is assumed to be an independent and identically distributed process, and hence the number of transmissions per packet has a geometric distribution. If successive packets were lost independently with probability equal to the average packet error rate of the channel, the assumption would be accurate. However, packet losses generally occur in bursts and the packet loss probability is usually variable and correlated.

Consider for example the traces in Fig. 9.2 taken from four distinct links in the Roofnet (the first trace was already given in Fig. 9.1 and the method of obtaining the traces was explained back there). Each of these four links has an ETX of approximately 2 during the testing period. Therefore, if ETX is taken as the metric for quality, these four links are identical. On the other hand, the sample variances of the delivery ratios are quite different for these links, i.e., these wireless links have similar long-term average behaviors, even though their short-term behaviors are quite different. Indeed, the sample coefficient of variation for the binary packet error sequences are 7.92, 2.16, 1.20 and 0.61.

One may ask whether it is possible to increase the frequency of ETX measurements and change the optimum paths accordingly more and more frequently until the "remaining" variability between updates is somewhat insignificant. Unfortunately, the update procedure involves significant amount of overhead in the network. If repeated frequently, it causes inefficient use of resources, extra interference and even instability of the routing algorithm. Therefore, the time-scale over which path-selection decisions are made is typically no less than tens or hundreds of packets; i.e., once a path between two nodes has been selected, it is likely to remain for several seconds. As shown in Fig. 9.2, there may be a huge channel variability over that time-scale and the ETX has to live with that.

In [15] Koksal *et al.* showed that the variability in short, as well as the longer time scales has a significant impact on the expected number of transmissions. It was

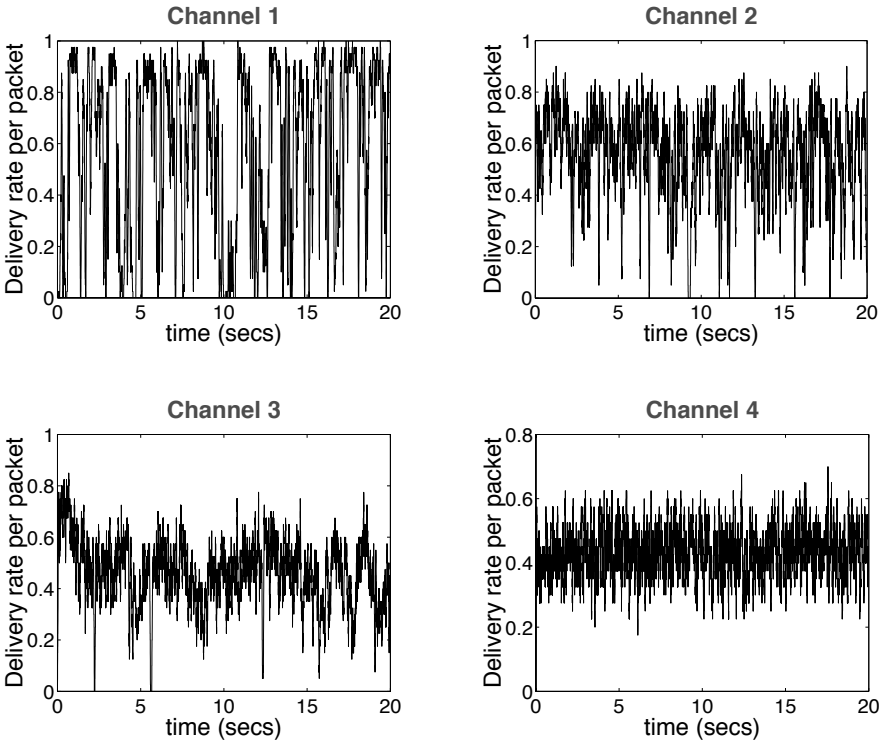



**Fig. 9.2.** Packet delivery rate for four distinct links of Roofnet.

further shown in [11] that, given two links in Roofnet, it not uncommon that the link with a lower ETX metric may in fact lead to a *higher* observed loss rate at the transport layer. The main reason for this is that good link-layer protocols do not try to retransmit lost packets forever but give up after a threshold number of attempts. When losses occur in bursts, picking the link in the middle of a burst-error situation would be bad *even* if it had a lower ETX.

To summarize, ETX can improve the throughput of a wireless mesh network by a significant amount compared to the hop count cost metric. However, ETX metric cannot track the variability of the channel at short time scales due to potential route instability.

## 9.3 Modified Expected Number of Transmissions (mETX)

This metric is built to overcome the shortcomings of ETX in the presence of channel variability. The development is based on a certain characterization of the channel

given in [15]. The authors developed tools to analyze the channels with non-iid losses and quantify the impact of channel variability on the number of transmissions. This lead to the mETX metric proposed in [11].

The model assumes that the bit error probability on a link is a (non-iid) stationary stochastic process. The variability of the link is modeled using the statistics of this stochastic process. Then, the mean number of transmissions is analytically calculated and the results show that it can be closely approximated with the first two order statistics of the bit error probability, summed over a packet duration. For mETX, the critical time scale for the link variability is the transmission time of a single packet including all its retransmissions.

The mETX metric is a function of the mean, $\mu_\Sigma$ and the variance, $\sigma^2_\Sigma$ of $\Sigma$, the bit error probability summed over a packet duration:

$$\text{mETX} = \exp\left(\mu_\Sigma + \frac{1}{2}\sigma^2_\Sigma\right). \tag{9.1}$$

The $\mu_\Sigma$ term represents the impact of slowly varying and static components in the channel (e.g., shadowing, slow fading), while the $\sigma^2_\Sigma$ represents the impact of relatively rapid channel variations (e.g., flat fading, interference) that the $\mu_\Sigma$ term (and hence the ETX) cannot track.

To estimate these two parameters, bit level information is necessary. Counting only the packet losses is not sufficient; thus, probe packets with a known content are used for estimation. The parameters $\mu_\Sigma$ and $\sigma^2_\Sigma$ are estimated by considering the number of errored bits in each probe packet. As in the ETX metric, each node sends probe packets periodically to calculate a loss rate sample and this information is passed to a moving average filter. Alternatively an EWMA filter can be used.

In [11], results of link measurements taken from 57 links that belong to distinct pairs of 12 different nodes of the Roofnet testbed were illustrated at a transmission rate of 11 Mbps. Based on the measurements, it was shown that the packet loss probability has a higher correlation coefficient ($\rho = 0.85$) with $\sigma^2_\Sigma$ than it has with $\mu_\Sigma$ ($\rho = 0.59$). Consequently, the link variability can be even more relevant than the ETX for packet losses. Also, by combining the impact of variability and the average loss rate, mETX achieves a drop of between 7%-50% in the average network loss rate (corresponds to an improvement of up to 60% in TCP throughput). The amount of reduction varies with the number of nodes and the node density.

The main drawback of the mETX metric is the complexity of the channel estimation. Firstly, the probe packets need to be processed at the bit level. This may not necessarily be an issue for the mesh networks due to the relative abundance of processing power, however, may be problematic for other platforms such as sensor networks. Secondly, the variance component, $\sigma^2_\Sigma$ increases with increased estimation error. Namely, a link may have a high mETX metric due to not only the high channel variability, but also the estimation error. Consequently, a better link with a high estimation error may end up having a higher metric than a worse link. On the other hand, one can justify the fact that links with more degraded information are less preferable, using the famous quote: “the shortest way home is the way you know.”

In the same way as ETX, the mETX can be adapted easily for radios that provide bit rate adaptation by normalizing the metric with respect to the transmission rate.

## 9.4 Effective Number of Transmissions (ENT)

The motivation for the ENT metric is to find routes that satisfy certain higher-layer protocol requirements. The challenge is finding a path that achieves high network capacity while ensuring that the end-to-end packet loss rate visible to higher layers (such as TCP) does not exceed a specified value. Given a loss constraint, picking the path that maximizes the link layer throughput may not be sufficient, because it may involve links with high loss rates. Because link-layer protocols give up after a certain threshold number of retransmissions ($M$), ETX and mETX may pick links that violate the loss rate requirement visible to higher layers. The ENT metric is designed to meet the desired goal.

Similar to the mETX metric, the ENT metric also characterizes the probability of bit error as a stationary stochastic process. Using a large deviations approach, it was shown in [11] that the probability of a packet loss (i.e., number of transmissions exceeding $M$) can be well approximated with

$$P_{\text{loss}} \approx \exp\left[-\frac{1}{2}\left(\frac{\log M - \mu_\Sigma}{\sigma_\Sigma}\right)^2\right] \tag{9.2}$$

for large packet sizes and large values of $M$. Now suppose the desired loss probability is $P_{\text{desired}}$ and let $\delta = -\log P_{\text{desired}}/\log M$. There is a one-to-one correspondence between the desired loss rate and $\delta$. Thus the parameter $\delta$ uniquely specifies $P_{\text{desired}}$. Note that $\delta$ is referred to as the *temporal diversity gain* in wireless communication. For a given $P_{\text{desired}}$ (i.e., $\delta$) to be met, $P_{\text{loss}} \leqslant P_{\text{desired}}$ and consequently,

$$\mu_\Sigma + 2\delta\sigma_\Sigma^2 \leqslant \log M. \tag{9.3}$$

The sum in the left side of (9.3) is defined as the log effective number of transmissions (i.e., $\log$ ENT) of the link.

One way to interpret (9.3) is as follows. Suppose the higher layer does not specify any loss probability constraint, i.e., $\delta = 0$. Condition (9.3) turns into a comparison of $\mu_\Sigma$ (hence the average bit error probability of the channel) with $M$. Thus, the higher-layer requirement turns into a condition involving average link parameters only, as is the case with ETX. Now suppose the higher-layer has a loss rate requirement, i.e., $\delta > 0$. In that case one needs to *underbook* the resources to meet the loss probability target. The amount of spare ETX that has to be put aside in order to accommodate channel fluctuations is $2\delta\sigma_\Sigma^2$. This margin allows the packet loss probability target to be met. As expected, this amount is directly related to the variability, $\sigma_\Sigma^2$, of the channel and the strictness, $\delta$, of the loss rate requirement. This interpretation of ENT is analogous to the notion of *effective bandwidth*, which was developed to model variable traffic sources in queueing networks. Indeed, ENT can be interpreted as the effective bandwidth of the discrete stochastic process, the number of transmissions.

ENT has a structure similar to mETX. The main difference is the extra degree of freedom due to the factor $2\delta$. Indeed, the mETX is the ENT evaluated at $\delta = 1/4$. Similar to the mETX, a by-product of ENT is to reduce the packet loss ratio observed by higher-layer protocols, *after* any link-layer retransmissions are done. Also, since exactly the same parameters are used in the ENT as in the mETX, the channel estimation procedure is identical.

On the other hand, the ENT metric is not additive as the ETX or the mETX. The metric is composed over successive links using minimax type routing algorithms. More precisely, among all the paths between two nodes, the path along which the links minimizes the maximum ENT is selected as the best route. Another algorithm that combines the ETX and the ENT metrics was proposed in [11]:

"For each link, compute its $\log$ ENT. Compare against $\log M$. Assign a cost of $\infty$ to the links that have $\log$ ENT $> \log M$ and assign a cost of ETX to the others. Between any pair of nodes use the path that minimizes the total cost." This algorithm focuses only on the feasible links, i.e., the ones that satisfy the application loss requirement, $P_{\text{loss}}$. It picks those with the minimum ETX among those.

The average network loss rate is also simulated with the link-level data acquired from the Roofnet. The set of feasible links are defined to be those that have an ENT of less than 16 for the $\delta$ parameter varied between 1 and 2.5. There were some interesting trends. First, the observed loss rates can be controlled by merely adjusting the "space parameter" $\delta$, which acts as a knob to control the performance. Not only it is guaranteed that each link has no more than a certain desired loss rate, but also the average network loss rate can be reduced by an amount between 7-55% depending on the network size and the control parameter $\delta$.

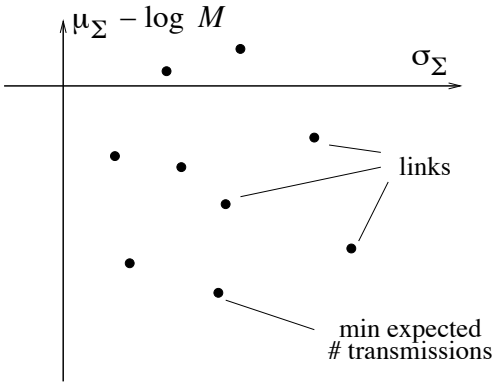
There is a catch, though. The loss rate does not decrease monotonically with increasing $\delta$. Beyond a certain threshold, the loss rate starts to increase. The reason for this transition is that too many links are eliminated for violating the loss constraint. Consequently, even many "decent" links are gone and no feasible paths remain between some node pairs and the network becomes disconnected.

Another benefit of ENT is that it can be calibrated. A network architect can adjust the $\delta$ parameter until the desired network performance is achieved. Indeed, the derivations in [11] are based on certain assumptions, which can be partly violated in different platforms and environments. It is useful to have a degree of freedom for the necessary adjustments.

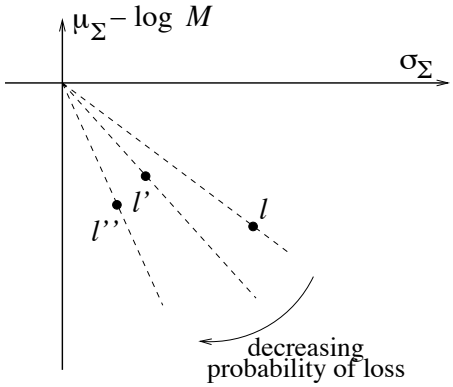
The main drawback of the mETX metric is valid for the ENT as well. Since the same channel estimation procedure is followed, the estimation error affects the ENT metric similar to the mETX metric.

## 9.5 Geometric Interpretation of Routing Cost Metrics

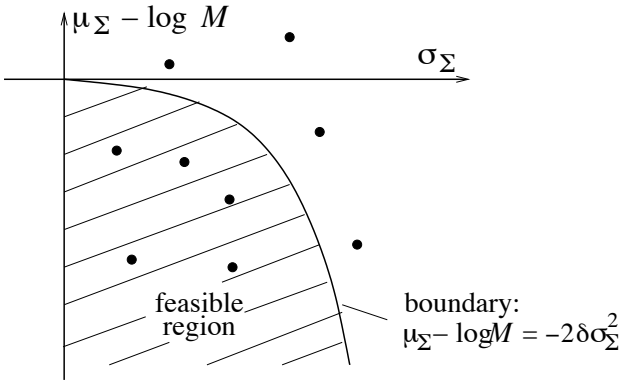
This section provides a unified geometric framework that combines the mean and standard deviation of the bit error rate process to visually compare the (quantized) loss rate, ETX, mETX and the ENT metrics.




(a) Each point in $(\sigma_\Sigma, \mu_\Sigma)$ coordinate space represents a link.




(b) The slopes of the dashed lines are representatives of the loss probability.




(c) The points in the feasible region satisfy $P_{\text{loss}} \leqslant \exp(-\delta(\mu_\Sigma - \log M))$.

**Fig. 9.3.** The geometry of the channel parameters and the loss probability.

Let us represent a wireless link by two parameters, $\mu_\Sigma$ and $\sigma_\Sigma$. Each link corresponds to a point in the coordinate space $(\sigma_\Sigma, \mu_\Sigma)$ as illustrated in Fig. 3(a). In these graphs we use $\mu_\Sigma - \log M$ instead of $\mu_\Sigma$ as the y-axis. This only introduces a linear shift, but simplifies our discussions. In this space, the point with the lowest ordinate value is the one that minimizes the ETX. Such links will be preferred by routing algorithms that employ $\mu_\Sigma$ as the link cost metric (e.g., ETX).

For any given point, the slope of the line connecting the origin to that point is $(\mu_\Sigma - \log M)/\sigma_\Sigma$. Combining this with Eq. (9.2), points with smaller slopes, i.e., points with larger $|(\mu_\Sigma - \log M)/\sigma_\Sigma|$ have lower loss probabilities. For instance, in Fig. 3(b), link $l$ has a higher loss probability (and hence a higher Quantized Loss Rate) than link $l'$. If the objective is to minimize the probability of loss, then the path selection algorithm should choose points with large $|(\mu_\Sigma - \log M)/\sigma_\Sigma|$ ratios.

The set of points with a certain diversity gain, i.e., for a given $\delta$, the links that satisfy $\delta = -\log P_{\text{loss}}/\log M$ lie on a parabola as shown in Fig. 3(c). Thus, the points outside the shaded region have a lower diversity gain and fail to satisfy the required constraint for $P_{\text{loss}}$. The shaded region can therefore be regarded as a *feasible region*. Notice that for $\delta = 0$ (i.e., no loss-rate requirement) the feasible region is the entire fourth quadrant. The region shrinks as $\delta$ is increased since the boundary of the region becomes more and more concave. Hence, a smaller number of links become feasible. For instance, consider the routing algorithm, which minimizes the ETX subject to an ENT constraint. It should pick the links with small ordinate values among the points in the feasible region.

Similarly, the set of links with a constant mETX constitute a parabola in the coordinate space $(\sigma_\Sigma, \mu_\Sigma)$. Indeed, the set of points with mETX equal to the constant $c$ lie on the parabola specified by

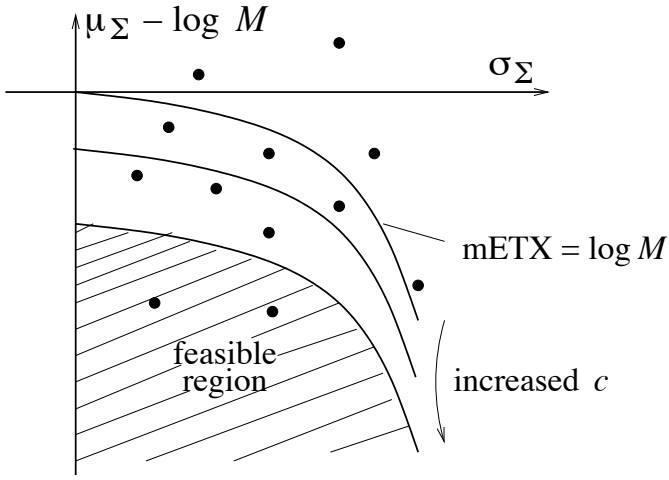
$$\mu_\Sigma + \frac{1}{2}\sigma_\Sigma^2 = c.$$

These parabolas can also be viewed as the boundaries for a feasible region, where the feasible links are those with mETX less than some given $c$ value. Constant mETX curves are illustrated in Fig. 4(a). As $c$ is reduced, the boundary moves farther away from the x-axis and consequently the set of points with smaller mETX shrink.

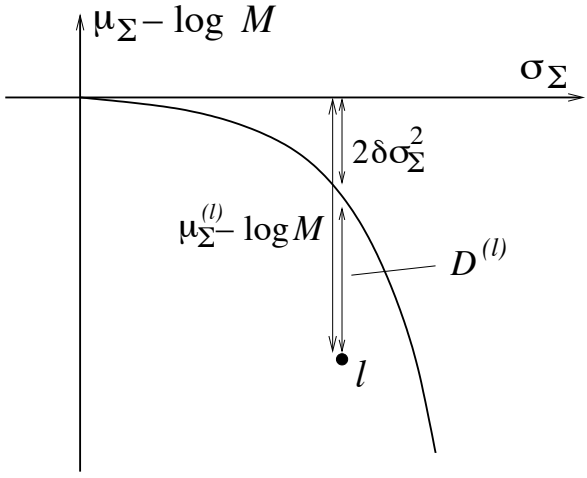
Finally, consider the vertical distance, $D^{(l)}$, between any admissible link $l$ : $(\sigma_\Sigma, \mu_\Sigma)$ and the boundary of a feasible region of links. As illustrated in Fig. 4(b),

$$\begin{aligned} D^{(l)} &= -(\mu_\Sigma - \log M) - 2\delta\sigma_\Sigma^2 \\ &= \log M - \text{ENT}^{(l)}(\delta). \end{aligned} \tag{9.4}$$

Hence, the feasible link that maximizes the vertical distance to the boundary of the feasible region is the one that minimizes the ENT. This means that, given an increase in the expected number of transmissions, the link with a small ENT is more likely to remain in the admissible region. Thus, if the objective is robustness with respect to the uncertainty in the measured parameters and to changes in the expected number of transmissions, the routing algorithm should choose points with smaller ENT.




(a) The set of points with mETX=c form the parabola $\mu_\Sigma + \frac{1}{2}\sigma_\Sigma^2 = c$. The points with smaller mETX lie inside the parabola.




(b) The vertical distance between the link $l$ and the boundary of the feasible region is $D^{(l)} = \log M - \log \text{ENT}$.

**Fig. 9.4.** The geometry of the mETX and the ENT.

## Conclusion

In this chapter we have studied seven routing cost metrics to be used for selecting good paths in wireless mesh networks. The following table summarizes these metrics, their benefits and drawbacks.

| Metric | Definition | Benefit | Drawback |
|---|---|---|---|
| *Hop Count* | # Hops | Simplicity | Chooses poor links |
| *Per hop RTT* | Delay/hop | Incorp. multiple factors | Self interference |
| *PktPair* | Transmit delay/hop | Reduces self interference | High overhead |
| *Loss Rate* | Packet loss rate | Eliminates lossy links | Low bandwidth paths |
| *ETX* | # Transmissions | Improves throughput | Fails under variability |
| *mETX* | ETX w/variability | Works w/variable links | Sensitive to est. errors |
| *ENT* | Effective bandwidth of link | Provides controlled QoS | Not composable solely |

There are a couple of directions along which the routing metrics can be studied further. One paradigm for routing in wireless networks is cooperative diversity. Cooperative diversity takes advantage of broadcast transmission to send information through multiple relays concurrently. Similar to the traditional routing protocols, cooperative schemes also require a differentiation mechanism among different links, which makes the cost metrics necessary. For instance in ExOR [16], once a packet is transmitted over a hop, it may be decoded correctly by a number of other nodes as well as the intended next hop. After the transmission, a priority ordering of such nodes is made to decide who will relay the packet next. This ordering is based on the total cost of different paths from each node that has a copy of the packet to the ultimate destination.

New metrics can be engineered specifically for cooperative communication as well as the multipath routing setting, in which data between a pair of nodes can be carried over multiple paths simultaneously. In the multipath scenario, the composability of a metric becomes critical not only along each path, but also over parallel paths.

Another extension can be to build metrics based on physical layer parameters such as the signal-to-noise ratio (SNR). This can shed further light on the impact of physical layer on optimal routing decisions. It can also bring the channel estimation in line with the common practice of using physical layer pilot symbols to estimate the channel gain.

Finally, studying the impact of adaptive coding and power control on routing is critical for WiFi and 802.11 based networks, since rate adaptation is an integral part of these standards.

## References

1. C. E. Perkins and P. Bhagwat, "Destination sequenced distance vector routing," in *Proc. of ACM SIGCOMM*, 1994.

2. M. Gerla, X. Hong, and G. Pei, "Fisheye state routing protocol for ad hoc networks," *Internet Draft, draft-ietf-manet-fsr-03.txt*, 2002.
3. P. Jacquet, P. Muhlethaler, A. Qayyum, Laouiti A, L. Viennot, and T. Clausen, "Optimized link state routing protocol," *Internet Draft, draft-ietf-manet-olsr-07.txt*, 2002.
4. C. E. Perkins and E. M. Royer, "Ad hoc on demand distance vector routing," in *Proc. of IEEE WMCSA*, 1999.
5. V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *Proc. of IEEE Infocom*, 1997.
6. D. B. Johnson, D. A. Maltz, Y. Hu, and J. G. Jetcheva, "The dynamic source routing protocol for Mobile ad hoc networks (DSR)," *Internet Draft, draft-ietf-manet-dsr-07.txt*, 2002.
7. J. C. Bicket, "Bit-rate selection in wireless networks," M.S. thesis, MIT, Feb 2005.
8. A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "A multi-radio unification protocol for IEEE 802.11 wireless networks," Tech. Rep. MSR-TR-2003-44, Microsoft Research, July 2003.
9. M. Yarvis, W. Conner, L. Krishnamurthy, J. Chhabra, B. Elliott, and A. Mainwaring, "Real world experiences with an interactive ad-hoc sensor network," in *Proc. of the International Workshop on Ad Hoc Networking*, August 2002.
10. D. DeCouto, D. Aguayo, J. Bicket, and R. Morris, "A high throughput path metric for multi-hop wireless routing," in *Proc. of ACM MobiCom*, 2003.
11. C. E. Koksal and H. Balakrishnan, "Quality aware routing metrics for time-varying wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 1984–1994, 2006.
12. D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris, "Link-level measurements from an 802.11b mesh network," in *Proc. of ACM SIGCOMM*, September 2004.
13. R. Draves, J. Padhye, and R. Patra, "Comparison of routing metrics for static multi-hop wireless networks," in *Proc. of ACM SIGCOMM*, 2004.
14. D. DeCouto, *High-Throughput Routing for Multi-Hop Wireless Networks*, Ph.D. thesis, MIT, June 2004.
15. C. E. Koksal, K. Jamieson, E. Telatar, and P. Thiran, "Impacts of channel variability on link-level throughput in wireless networks," in *Proc. of ACM SIGMETRICS/Performance*, 2006.
16. S. Biswas and R. Morris, "ExOR: Opportunistic multi-hop routing for wireless networks," in *Proc. of ACM SIGCOMM*, August 2005.

# 10

# Cross-layer Solutions for Traffic Forwarding in Mesh Networks*

V. Baiamonte, C. Casetti, C. F. Chiasserini, and M. Fiore

Politecnico di Torino, Italy
{valeria,claudio,carla,marco}@tlc.polito.it

## 10.1 Introduction

Wireless mesh networks [1] is an emerging wireless technology that allows robust and reliable wireless broadband service access at relatively low cost. They include two types of nodes: mesh routers and mesh clients. Both types of nodes operate not only as hosts but also as routers, forwarding packets on behalf of other nodes that may not be within direct wireless transmission range of their destinations; in addition, a mesh router may have gateway/bridge functionalities [1]. Wireless mesh nodes dynamically self-organize and self-configure, automatically establishing and maintaining mesh connectivity among themselves.

This chapter focuses on a mesh network using the IEEE 802.11 technology. Consider the network section including mesh clients and routers (hereinafter also called nodes) that wish to connect to a mesh gateway, through either direct or multihop communications. The problem addressed here is how to transfer traffic between the wireless nodes and a gateway in a fair, efficient manner.

Routing protocols for wireless networks are usually designed considering that all nodes within transmission range of a transmitter are equivalent. However, this is often false, as the quality of the channel toward (and from) different one hop neighbors may significantly vary with distance, presence of obstacles and interfering transmissions. Also, since 802.11 off-the-shelf devices implement rate-adaptation techniques, the link quality directly determines the data transmission rate to be used between pairs of nodes. Thus, measuring routing distances in terms of number of hops may be misleading, as routing through a larger number of high-rate hops may lead to a higher network throughput with respect to performing fewer low-rate forwards [2,3]. An example of this behavior can be observed in the simple scenario of Fig. 10.1. Here, all nodes are within receive range of each other, and both $A$ and $B$ send CBR over UDP data to $C$ using the 802.11 Distributed Coordination Function (DCF) to access the radio channel. Assume that $A$ enjoys an optimal channel quality toward $C$,

*This work was supported by the Italian Ministry of University and Research through the TWELVE and the MEADOW projects.

while $B$ experiences a low quality channel due to distance and/or obstacles. The resulting decrease in $B$'s data transmission rate is the cause of the well-known 802.11 anomaly phenomenon [4], which reduces the overall network throughput, as shown in Fig. 10.2. However, if the channel quality between $B$ and $A$ is good enough and $A$ relays data from $B$ toward $C$, then the resulting overall network throughput is better than that obtained with a 2 Mb/s direct data transfer between $B$ and $C$. Even higher improvements are obtained when employing TCP, which introduces reverse acknowledgments flows from $C$ to $A$ and $B$.

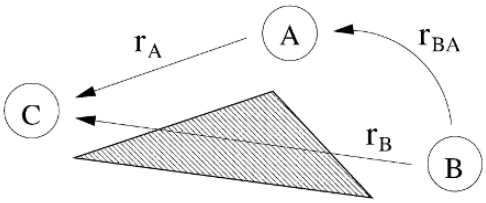



**Fig. 10.1.** Single-hop anomaly scenario.

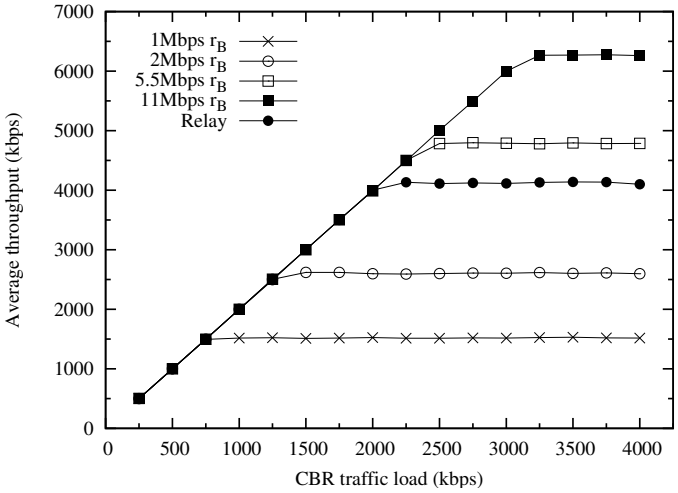



**Fig. 10.2.** Network throughput versus per-node offered load, with $r_{AC}$ and $r_{BA}$ equal to 11 Mbps and varying $r_{BC}$.

An idea to overcome the problem described above is to design a routing protocol accounting for medium access control (MAC) and physical layer performance in the route computation, by making multiple fast hops preferable to single slow ones. The joint design of MAC and routing schemes is however a challenging issue. Indeed, even if the use of relay nodes may alleviate the anomaly effect and increase the

system throughput, as shown in the example above, the role of relay is a thankless one: in addition to its own traffic, the relay node must carry other nodes' traffic. Therefore, some incentives are needed so that the throughput of the relay node is close to that of the same node without relay traffic.

In this chapter, we first define and examine two relay strategies that aim at giving relay nodes some incentives for their roles, while at the same time enhancing the overall throughput of the network. The first strategy involves the implementation of two Logical Link Control (LLC) queues at each node: one handles 'local' traffic, the other collects relay traffic from other nodes. The second technique relies on the enhanced distributed channel access (EDCA) specified by the IEEE 802.11e draft standard [5]. We then use experimental measurements and simulation results to derive some guidelines on designing an efficient, cross-layer, relay selection scheme that accounts for quality and transmission rate of the available links. We describe a relay selection algorithm and define a relay-quality aware routing [6], as an extension of the Optimized Link State Routing (OLSR) [7] protocol.

The rest of the chapter is organized as follows. Section 10.2 reviews some work on routing in ad hoc networks. The benefits of multihop forwarding in counteracting the anomaly effect are discussed in Section 10.3, with the help of experimental measurements. Section 10.4 describes the proposed traffic forwarding strategies and their performance. Section 10.5 summarizes the major lessons learned from experimental measurements and simulation results, and presents our relay selection algorithm. Section 10.6 describes its implementation and shows some performance results.

## 10.2 Packet Relaying in Wireless Networks

Routing in wireless networks has received a great deal of interest, and several schemes have been proposed which exploit various metrics for route selection.

In [8] it has been observed that a routing scheme using the hop count as a metric for route selection may not be the best choice. Indeed, while this scheme may be appropriate in single-rate networks, in a multi-rate environment it tends to select short paths composed of maximum length links. Since long distance links operate at low rates, poor throughput performances are likely to be obtained. To select high-throughput paths in multihop networks, the use of the expected transmission count (ETX) metric is proposed in [2]. Based on the ETX metric, the route featuring the fewest expected number of transmissions (including retransmissions) to deliver a packet is chosen.

The solutions presented in [3, 9, 10] design multirate-aware routing schemes to increase utilization of 802.11-based, multihop networks. In particular, the key idea in [9] is to change the next-hop node to another node where higher data rates are available. In [3], the authors proposed a scheme for route selection that attempts at solving the 'anomaly' effect thus yielding improved throughput performance. Whenever a source node requires a route to a destination node, the proposed scheme determines the best path based on the collision probability at the MAC layer and the

**Table 10.1.** Testbed details: Hardware and software tools.

| | |
|---|---|
| NIC | Cisco wireless 802.11a/b/g chipset Atheros |
| NIC Driver | Madwifi |
| OS | Linux, live CD Slackware based |
| PHY | 802.11b |
| Enabled Interfaces | Monitor mode/Ad-Hoc Mode |

available bandwidth. The work in [10] presented a modified routing metric with respect to ETX, which accounts also for the bandwidth available at the 802.11 nodes.

Similar issues were addressed in [11], where a MAC-layer, relay-enabled Point Coordination Function (PCF) protocol was presented. The scheme allows packet delivery through a relay node if the direct link has low quality and low rate.

Finally, in [12] an analytical tool was presented, which evaluates the expected throughput along a route in 802.11 multihop mobile networks.

## 10.3 Preliminaries

Given an 802.11 network with infrastructure where communication nodes use the DCF scheme, the overall system throughput significantly decreases in presence of low data-rate transmitters. The reason for this behavior is that DCF is based on the CSMA/CA which provides an equal, long-term channel access probability to all nodes. When a low data-rate node seizes the channel, it keeps it for a long time, thus penalizing other high-rate nodes [4].

In this section, we describe the experimental measurements performed to study the 802.11 anomaly and evaluate the advantages of the multihop relay mechanism.

### 10.3.1 Experimental Measurements

Experiments have been carried out in outdoor environments with the hardware and software equipment described in Table 10.1. We employ three different laptops behaving as wireless ad hoc nodes. Every laptop is equipped with a WLAN PCMCIA network interface card (NIC). Thanks to the functionalities provided by the MadWifi driver, we can enable multiple *virtual network interfaces* on each WLAN network interface card: we exploit this feature to create on each machine two virtual interfaces, operating concurrently in *ad hoc mode* and in *monitor mode*, respectively. The former mode allows every node to work as a router; the latter is used to dump traffic traces at each node. Monitor mode is indeed necessary to collect all packets heard on the channel, and to send them to the upper layers to dump them on a log file. For our measurements, we use packet-level traces that are collected at every interface involved, with the aim of having a faithful observation of the channel activity from any possible active radio transceiver.

The traffic generator *iperf* [13] is employed to generate both UDP and TCP traffic, although here we show results under the UDP traffic scenario only. UDP data flows consist of CBR data. The software used for traffic sniffing is called *Tethereal* – the command-line version of the more known *Ethereal*.

The two network configurations that we test are as follows. A wireless node, WN1, receives traffic from sources WN2 and WN3. WN2 uses a direct link at 11 Mb/s toward WN1 (flow tagged as Flow 2-1); WN3 can use either a direct link with data rate set at 1 or 2 Mb/s (tagged Flow 3-1), or a single-relay transmission by sending its traffic to WN2 through a 11 Mb/s link (flows tagged as Flow 3-2 and Flow 2-1 R, respectively).

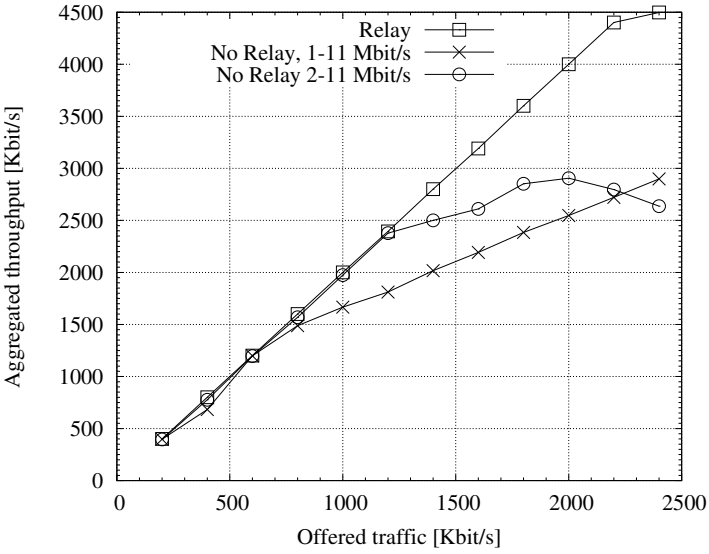



**Fig. 10.3.** Aggregate throughput under the UDP traffic scenario.

## Results

Fig. 10.3 presents the aggregate throughput as a function of the traffic generated at one single node. Each point on the graph represents the average throughput over 4 different experiments with identical settings. Every single experiment spanned over a 30 seconds time interval. The aggregate throughput is obtained as the sum of the throughput gained by the two source nodes (WN2 and WN3). We considered the traces collected at the receiving node, i.e., WN1, and took into account all packets correctly delivered and acknowledged.

In each plot, three different scenarios are compared: the Relay configuration, the configuration with No Relay where WN3 transmits at 1 Mb/s and the configuration

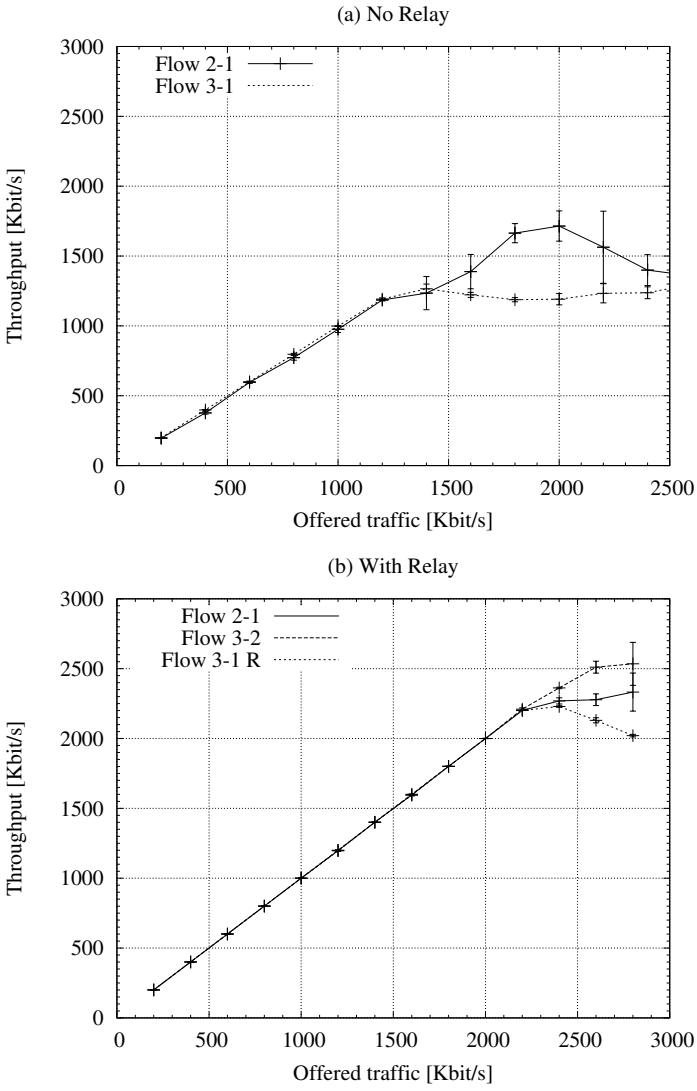



**Fig. 10.4.** Single flows throughput under the UDP traffic scenario.

with No Relay where WN3 transmits at 2 Mb/s. As expected, Fig. 10.4 shows that the more the offered traffic, the higher the total throughput achieved by the network, resulting in a linear slope till the network saturation is reached. It is interesting to notice the advantage of the single-relay forwarding against the slower direct link: the relay configuration clearly allows higher values of the aggregate throughput with respect to the other two scenarios. The network with relay starts being saturated when each wireless node generates 2.4 Mb/s, while with the direct links at 1 Mb/s and 2 Mb/s saturation is reached when the offered traffic per source is 600 kb/s and 1400 Kb/s, respectively.

The throughput achieved by each traffic flow is depicted in Figs. 10.4(a) and 10.4(b) for the No Relay and the Relay scenario, respectively, under UDP traffic. The vertical continuous lines represent the computed confidence intervals. In Fig. 10.4(a) the penalized node, WN3, is transmitting directly to the destination at 2 Mb/s. The performance anomaly can be clearly detected here. When saturation is reached (i.e., the offered traffic is 1400 Kb/s), not only the throughput of Flow 3-1 is decreased (dotted line), but also the throughput of the fast node (bold line) is significantly reduced, almost reaching the same values as the slow node for high traffic load. When the relay node is employed (Fig. 10.4(b)), we obtain much better performance, and the throughput starts decreasing only when the traffic sources generate 2.4 Mb/s each. Interestingly, for high network load, the flow achieving the highest throughput is Flow 3-2, while the slowest flow is the relayed traffic (Flow 2-1 R). The reason for this behavior is that WN2 has to relay traffic besides transmitting its own, its throughput is therefore lower than the one of Flow 3-2. As for Flow 2-1 R, several packets belonging to relayed traffic are dropped at WN2 due to buffer overflow, leading to a reduction in the flow performance. Results on loss probability and retransmission probability confirmed the above considerations.

## 10.4 Forwarding Strategies

Here we present our forwarding strategies, which aim at providing an efficient as well as fair traffic delivery in mesh networks. In the following we consider the traffic to flow in the downlink direction, i.e., from a mesh gateway to wireless mesh nodes. Similar observations are valid for the uplink traffic.

### 10.4.1 The Split Queues Approach

This approach provides for two LLC queues for each MAC queue at the wireless node. We will call these queues “Local” and “Relay” queues: the former collects packets originated locally and the latter collects relay traffic from other nodes. A dispatcher at the queues input reads network-layer packets as they are handed down to the link layer, and, based on the network-layer address, feeds them to the corresponding queue. A scheduler at the queues output serves the Local or the Relay queue according to a simple Split Queues (SQ) algorithm that operates in the following fashion:

- the Relay queue is served if the Local queue was served in the previous round, OR if the Local queue size is smaller than a threshold $T_l$;
- the Local queue is served if the Relay queue was served in the previous round AND if the Local queue size is larger or equal than a threshold $T_l$.

The algorithm is work-conserving, so if either queue is empty, the other is served by default. This solution can be implemented on 802.11b wireless cards, requiring only a modification of the LLC driver and no hardware changes.

### 10.4.2 The Access Category Approach

The second approach that we evaluate relies on 802.11e EDCA capabilities. In the following, we first describe the main features of EDCA in order to make the paper self-contained, then we introduce the proposed strategy.

*Overview of 802.11e EDCA*

Like 802.11 DCF, EDCA is based on the CSMA/CA scheme and employs the concepts of Inter Frame Space (IFS) and backoff to distributively control the channel access; furthermore it introduces the following innovations.

- When an 802.11e node seizes the channel, it is entitled to transmit one or more packets for a time interval named Transmission Opportunity (TXOP); a TXOP is characterized by a maximum duration, called TXOP_Limit.
- Various Access Categories (ACs) are defined, each of which corresponds to a different priority level and to a different set of parameters to be used for contending the channel. In particular, an 802.11e node operating under the EDCA function includes up to four MAC queues; each queue corresponds to a different AC and represents a separate instance of the CSMA/CA protocol. A queue employs the following parameters to access the channel: (i) the Arbitration Inter Frame Spacing (AIFS[AC]), similar to the DIFS used in DCF, (ii) the Minimum and the Maximum Contention Window ($\mathrm{CW_{min}}$[AC], $\mathrm{CW_{max}}$[AC]), (iii) and the TXOP_Limit[AC]. The higher the AC priority is, the smaller the AIFS[AC], $\mathrm{CW_{min}}$[AC] and $\mathrm{CW_{max}}$[AC] are. The larger the TXOP_Limit[AC], the greater the share of capacity of the AC. However, the values of $\mathrm{CW_{min}}$[AC] and $\mathrm{CW_{max}}$[AC] have to be carefully chosen so as to avoid high collision probability among traffic flows belonging to the same AC, and the value of AIFS must be at least as long as the DIFS interval (the only exception is for the AP that can use a shorter AIFS in order to gain control of the channel for coordination purposes, such as beacon transmission).
- Within every 802.11e node, a scheduler solves *virtual collisions* among the AC queues, i.e., among the various CSMA/CA instances, by always enabling the queue associated with the highest priority to transmit.

*The Proposed Approach*

Consider that several queues are implemented at the MAC layer, each associated to an Access Category (AC). Referring for the sake of concreteness to an 802.11e node operating under the EDCA function, up to four MAC queues can be defined, each of which corresponds to a different priority level and to a different set of parameters to be used for contending the channel.

In the AC approach, packets are dispatched to a different MAC queue, hence to a different Access Category, based on the locality of their address, and the type of traffic. The choice that we have investigated performs the following assignments to Access Categories in order of decreasing priority:

- AC[0] - Local UDP
- AC[1] - Local TCP
- AC[2] - Relay UDP
- AC[3] - Relay TCP

It is assumed that all other nodes implement 802.11e as well, and that, if they are not acting as relays, their traffic can only enjoy AC[2] or AC[3] status, which means a lower priority than local traffic from relay nodes. This provides an incentive to users to have their wireless node act as a relay, since this will allow their own local traffic to receive high-priority access. As a further incentive, relay nodes may be allowed an extended burst of packet transmissions by tweaking their 802.11e TXOPs. In our study we have experimented with TXOPs that allowed relay nodes to transmit as many as three back-to-back packets. As can be easily seen, this solution requires an 802.11e-enabled wireless LAN, and, though results are more promising than the previous approach, its implementation depends on the availability of 802.11e technology.

### 10.4.3 Simulation Results

We study the performance of our strategies under various network scenarios, using the *ns-2* simulator. Firstly, we present two sets of results related to sample single-relay and multiple-relay configurations, deriving preliminary observations. These results provide a comparison of the performance of the different strategies in a generic setting, allowing the identification of the most promising solution. Next, we try to establish elementary patterns of behavior in a simple four-node configuration with different bit rates and varying number of relays. Our aim is to couple the best relaying approach with the most effective choice of relay position and number.

*Single- and Multiple-relay Scenario*

In the following, we examine the results obtained by assuming the network topologies in Figs. 10.5 and 10.6. In both configurations we have 5 ‘fast’ nodes, i.e., close to the mesh gateway (GW) and enjoying an 11 Mb/s link with the gateway. Then, there are 5 ‘slow’ nodes to which the gateway can transmit either through a direct

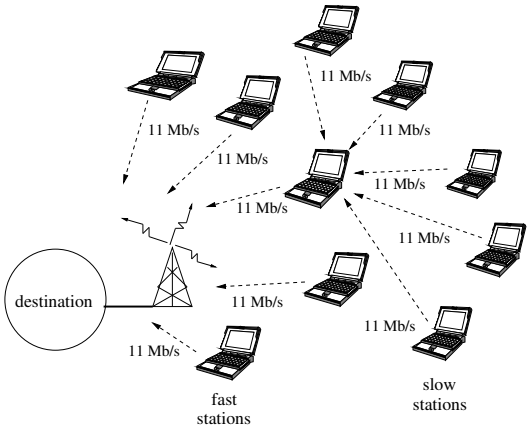



**Fig. 10.5.** Simulated topology: Single relay.

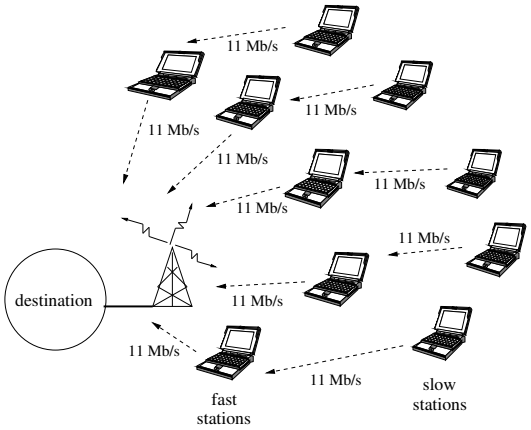



**Fig. 10.6.** Simulated topology: Multiple relays.

link at 1 Mb/s, or by using a 'fast' node as a relay. Note that relay nodes can transmit to slow nodes through an 11 Mb/s link, and vice versa. We consider both the case of a single relay for all slow nodes, and the case of multiple relays, one for each slow node.

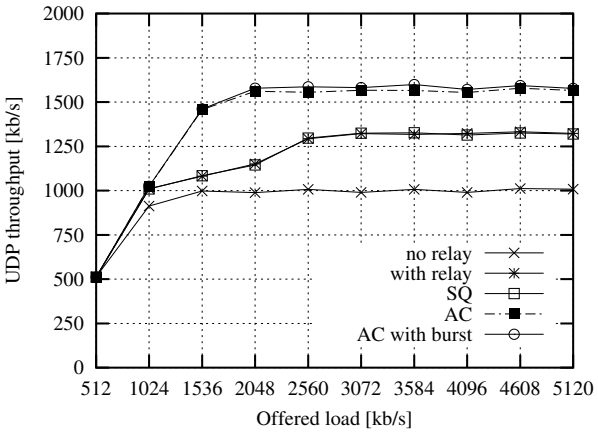



**Fig. 10.7.** Aggregate UDP throughput achieved by the different strategies: Single relay.

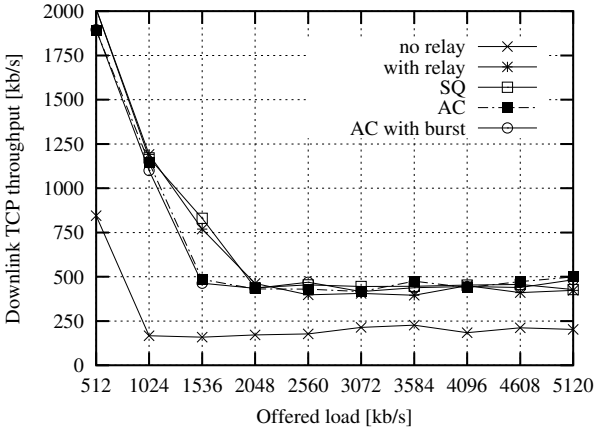



**Fig. 10.8.** Aggregate downlink TCP throughput achieved by the different strategies: Single relay.

Traffic in the simulations is a mix of downlink UDP flows (mimicking a video streaming application) and client-server interactions over TCP flows (with the wire-

less nodes acting as clients and the server being located on the wired portion of the network).

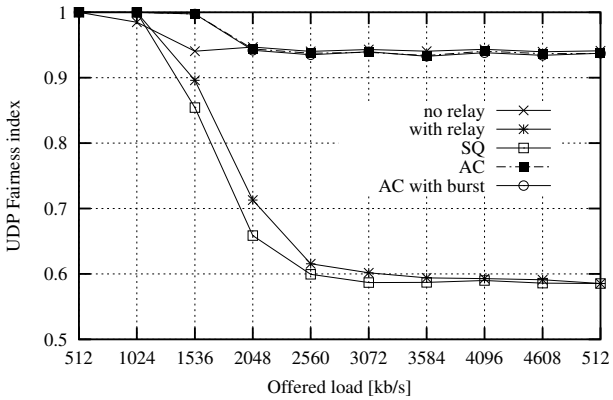



**Fig. 10.9.** UDP Throughput Fairness index for the different strategies: Single relay.

Figs. 10.7 and 10.8 show the aggregate throughput achieved by UDP and TCP downlink transfers, as a function of the offered UDP load, in the single-relay scenario. The benchmark results using standard configurations show that, while the "no relay" case clearly suffers from the anomaly effects of low-rate transmissions from far nodes, the use of a relay node (curves labeled "with relay") provides additional throughput for UDP and TCP flows. Compared to these benchmarks, the improvement introduced by our solution is clearly visible only for the Access Category approach (UDP throughput labeled "AC" and "AC with burst"). The use of the Split Queues ("SQ") only benefits the TCP throughput, while it is irrelevant as far as UDP is concerned. Indeed, it behaves similarly to the standard relay case. A further insight is gained by looking at the UDP throughput fairness index [14], that, for a variable $X$, is defined as,

$$F(X) = \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n \sum_{i=1}^{n} X_i} \tag{10.1}$$

where the $X_i$'s denote the $n$ samples of $X$. Fig. 10.9 shows that the SQ approach is the least fair (and a little less fair than the standard relay case); such poor performance highlights the inadequacy of the SQ approach at handling relay traffic when only one node acts as relay for multiple nodes.

If we now consider the multiple-relay scenario, in Figs. 10.10 and 10.11, the UDP throughput do not differ much from the single-relay ones, save for the AC approach using extended three-packet bursts, which reaches higher values, at the expense of fairness (relay nodes manage to transmit more local traffic). As far as fairness is concerned, the Split Queue approach evenly distributes the throughput among near and far nodes without incurring the penalizing unbalance experienced by the standard

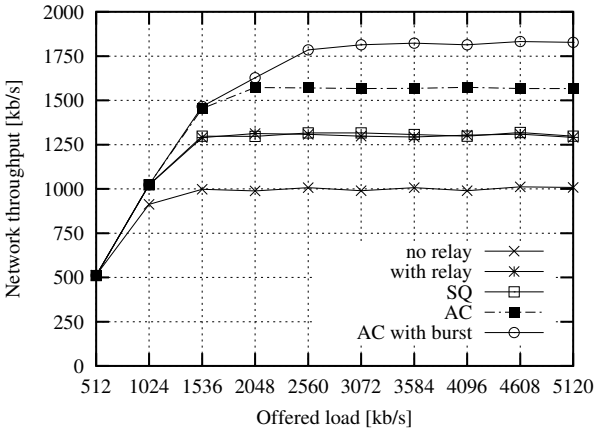



**Fig. 10.10.** Aggregate UDP throughput achieved by the different strategies: Multiple relays.

relay case, where far nodes achieve higher throughput than the local traffic of relay nodes.

To summarize, throughput and throughput fairness index results in the latter scenario provide us with a clear hierarchy of solutions. The Split Queue approach guarantees the same additional throughput of a simple relay solution, but can provide very high and load-insensitive fairness, which cannot be said of the simple relay solution. On the other hand, the use of 802.11e introduces remarkable throughput gains; the performance is even better if coupled with relay burst transmission. The fairness at best matches that of the no relay case, while it is worse for the burst case, as can be expected.

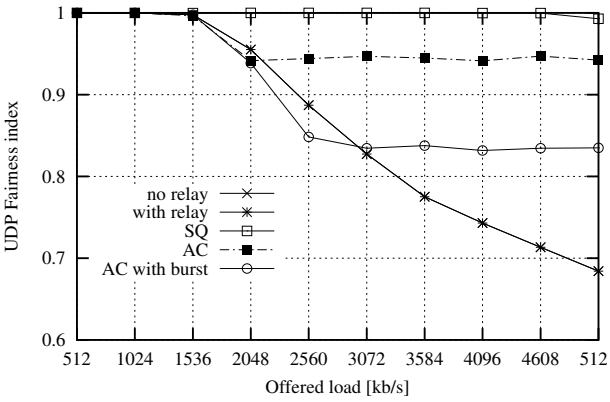



**Fig. 10.11.** UDP throughput fairness index for the different strategies: Multiple relays.

*Four-node Scenario*

Having ascertained that the Access Category approach provides remarkable performance gains in generic scenarios, we now seek to pinpoint the best combination of relay strategy and relay position (as well as number of relays). The availability of several bit rates across the network also provides us with an indication of the impact of the 802.11 performance anomaly due to a mixture of different rates. The network configuration shown in Fig. 10.12 allows us to single out test case behaviors without the superposition of effects that a crowded WLAN exhibits. The nodes are spaced so that the quality of the link between the mesh gateway and farther nodes is decreasing with distance, and nodes are within radio range of each other. If no node acts as relay, link speeds of nodes WN1 to WN4 are, respectively, 11, 5.5, 2 and 1 Mb/s, corresponding to the four bit rates allowed by the IEEE 802.11b standard. Table 10.2 also summarizes the configuration, whose results are referred to as the "no relay" case in the plots, and are used as reference case.

We only consider UDP downlink traffic (i.e., no background TCP traffic).

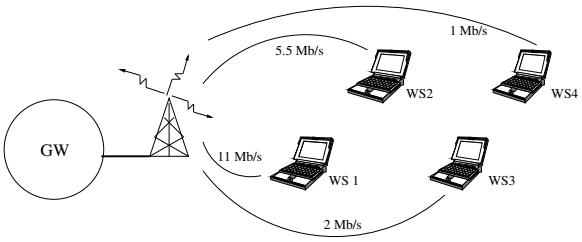



**Fig. 10.12.** Simulated topology: Four-node configuration.

We begin by comparing results for three different single-relay configurations. Tables 10.3, 10.4 and 10.5 show the relay position, marked by $(R)$ beside the node name. The "Next Hop" column indicates the destination of local and, possibly, relay traffic of every node in the *uplink* direction (downlink transfers are routed over the same path in the opposite direction); the "Bit Rate" column details the bit rate at which data are exchanged with the next hop.

**Table 10.2.** No relay.

| Node | Bit Rate (Mb/s) | Next Hop |
|---|---|---|
| WN1 | 11 | GW |
| WN2 | 5.5 | GW |
| WN3 | 2 | GW |
| WN4 | 1 | GW |

The comparison of throughput in Figs. 10.13, 10.15 and 10.17, corresponding to a relay positioned at WN3, WN2 and WN1, respectively, confirms that the best overall results are still yielded by the AC approach. However, only configuration 2 (WN2 as relay and lowest bit rate at 5.5 Mb/s) achieves a significant gain over the no relay case, mainly because the position of the relay guarantees that no low-rate transmissions take place, thus minimizing the impact of the anomaly. Configuration 3 (WN1 as relay) not only yields low throughput, comparable to the no relay case, but it is also hardly fair unless the AC approach is used (Fig. 10.18). Figs. 10.14 and 10.16 show that fairness is acceptable for configurations 1 and 2.

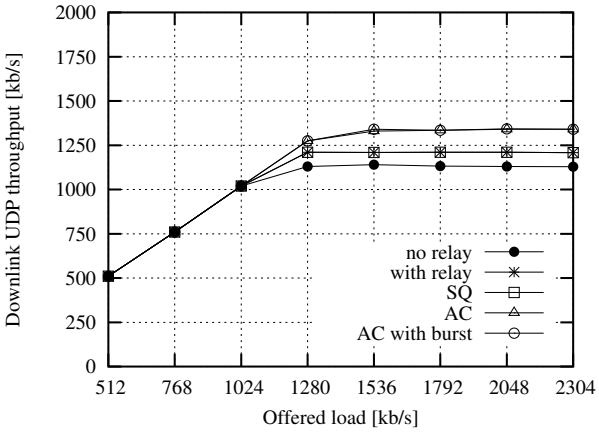



**Fig. 10.13.** Downlink UDP throughput with a single relay (conf. 1).

Although not shown here for lack of space, uplink UDP results exhibit similar behaviors in the three configurations considered.

The case of multiple relays has been studied in two sample cases: multiple relays supporting different nodes (thus introducing a single additional hop to communications between far nodes and gateway) and multiple recursive relays (introducing more than one hop between far nodes and gateway). The routing settings chosen by each configuration are summarized in Tables 10.6 and 10.7, respectively. The inspec-

**Table 10.3.** Single relay - configuration 1.

| Node | Bit Rate (Mb/s) | Next Hop |
|---|---|---|
| WN1 | 11 | GW |
| WN2 | 5.5 | GW |
| WN3 (R) | 2 | GW |
| WN4 | 11 | WN3 |

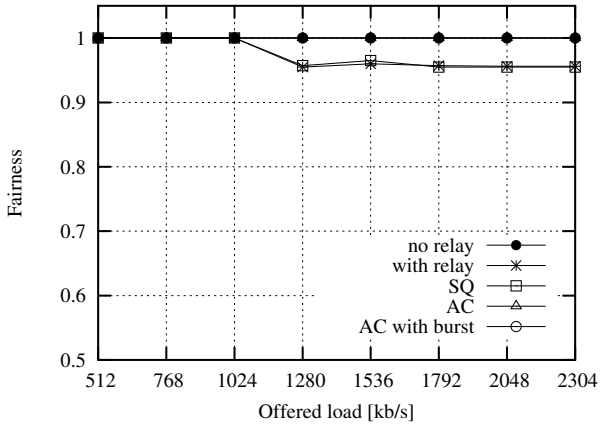



**Fig. 10.14.** Downlink UDP throughput fairness index for the different strategies: Single relay (conf. 1).

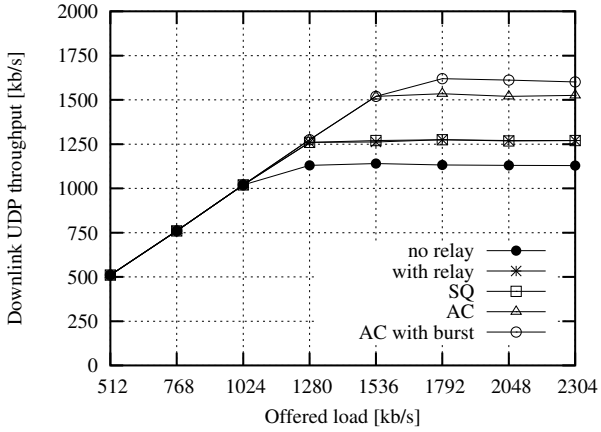



**Fig. 10.15.** Downlink UDP throughput with a single relay (conf. 2).

**Table 10.4.** Single relay - configuration 2.

| Node | Bit Rate (Mb/s) | Next Hop |
|---|---|---|
| WN1 | 11 | GW |
| WN2 (R) | 5.5 | GW |
| WN3 | 11 | WN2 |
| WN4 | 5.5 | WN2 |

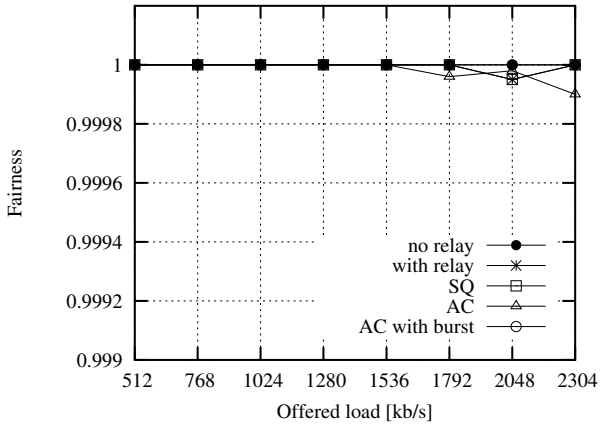



**Fig. 10.16.** Downlink UDP throughput fairness index for the different strategies: Single relay (conf. 2).

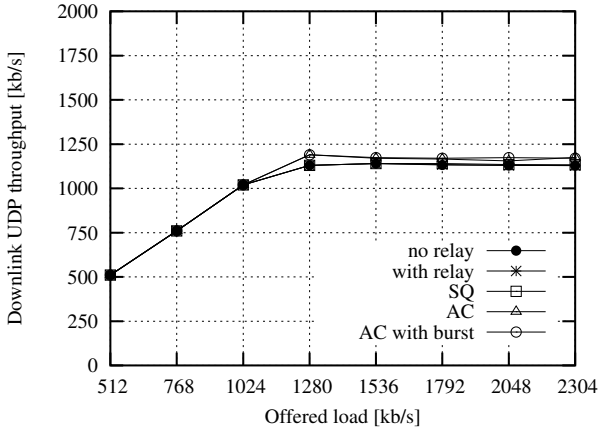



**Fig. 10.17.** Downlink UDP throughput with a single relay (conf. 3).

**Table 10.5.** Single relay - configuration 3.

| Node | Bit Rate (Mb/s) | Next Hop |
|---|---|---|
| WN1 (R) | 11 | GW |
| WN2 | 11 | WN1 |
| WN3 | 5.5 | WN1 |
| WN4 | 2 | WN1 |

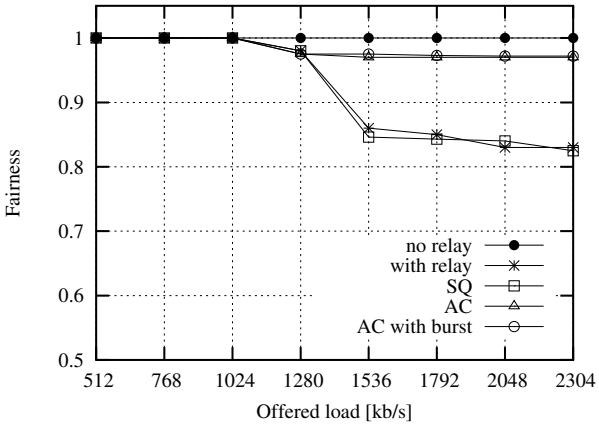



**Fig. 10.18.** Downlink UDP throughput fairness index for the different strategies: Single relay (conf. 3).

tion of results in Figs. 10.19, 10.20, 10.21, and 10.22 suggests that while multiple one-hop relays are beneficial to the system (even more so using our scheme), recursive relays are to be avoided. The dismal throughput performance highlighted in Fig. 10.21 can be ascribed to the offered load increase caused by copies of the same packet being transmitted several times across relays, though at high speed. Similarly, fairness (Fig. 10.22) is seriously compromised by the different load imposed to relay nodes (decreasing for nodes farther from the gateway).

**Table 10.6.** Multiple relays - configuration 1.

| Node | Bit Rate (Mb/s) | Next Hop |
|---|---|---|
| WN1 (R) | 11 | GW |
| WN2 (R) | 5.5 | GW |
| WN3 | 5.5 | WN1 |
| WN4 | 5.5 | WN2 |

**Table 10.7.** Multiple relays - configuration 2.

| Node | Bit Rate (Mb/s) | Next Hop |
|---|---|---|
| WN1 (R) | 11 | GW |
| WN2 (R) | 11 | WN1 |
| WN3 (R) | 11 | WN2 |
| WN4 (R) | 11 | WN3 |

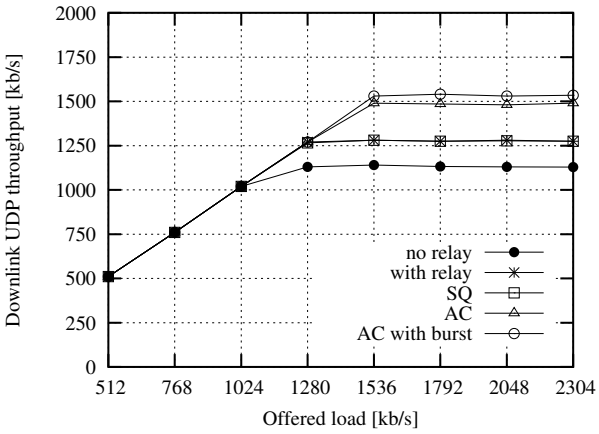



**Fig. 10.19.** Downlink UDP throughput with multiple relays (conf. 1).

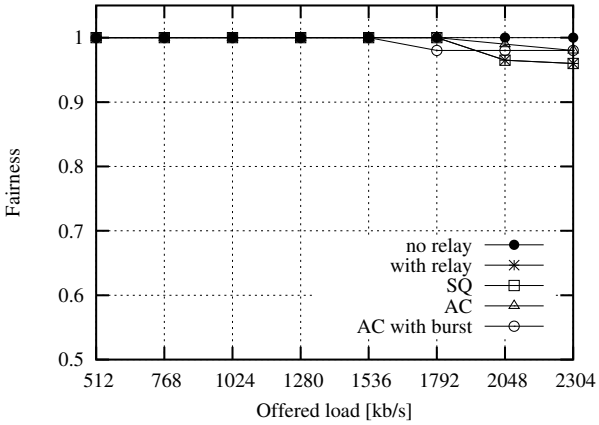



**Fig. 10.20.** Downlink UDP throughput fairness index for the different strategies: Multiple relays (conf. 1).

## 10.5 The Fair Relay Selection Algorithm

The results shown in the previous sections allow us to draw a set of observations that might be used as a springboard to design a Fair Relay Selection Algorithm (FRSA). Below, we summarize the main conclusions that simulation has provided:

- the 802.11 performance anomaly can be solved through the use of relays;
- using differentiation techniques to support relay and local traffic is beneficial;
- single relays offer a good compromise between the use of higher bit rates and lower numbers of packet replicas;
- best performance is achieved if relay(s) uses similar rates between gateway and mesh nodes;

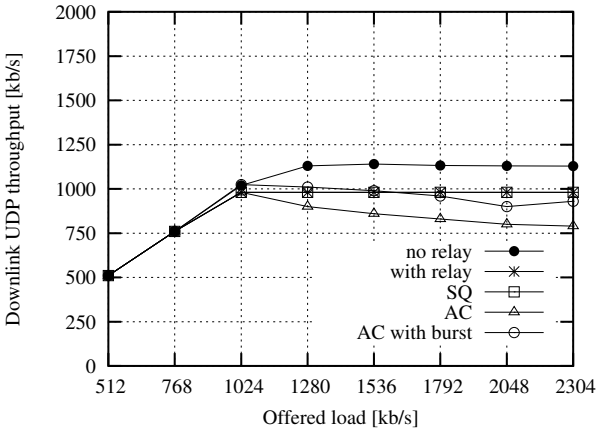



**Fig. 10.21.** Downlink UDP throughput with multiple relays (conf. 2).

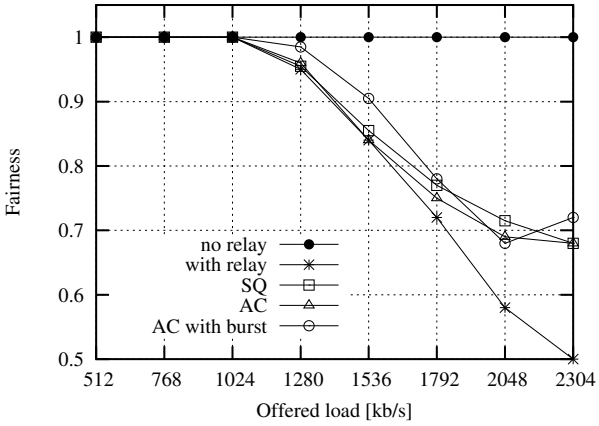



**Fig. 10.22.** Downlink UDP throughput fairness index for the different strategies: Multiple relays (conf. 2).

- if multiple relays are available, it is best to distribute far nodes among them;
- recursive relaying should be avoided.

The definition of FRSA leverages existing routing algorithms for ad hoc networks; we choose to refer to a proactive routing scheme, such as OLSR, since it seemed more suitable for including FRSA (see the following section). Also, given the limited support currently available for IEEE 802.11e capabilities in commercial hardware, we choose not to use any MAC-layer differentiation techniques. Nor do we implement the Split Queues approach, in order to test the benefits deriving from the new FRSA algorithm alone.

The main requirement of the algorithm is the identification and ordering of the available paths between a node and the gateway. The paths providing the best throughput will be ranked first.

Following common notation, the resulting topology information stored by each wireless node can then be mapped into a graph $G(V, A)$, where $V$ is a set of vertices, representing the network nodes ($n_i$), and $A$ is a set of arcs, representing links between pairs of nodes ($l_{ij}$); each link is associated to a cost $c_{ij}$, a function of the link feasible rate (i.e., the highest bit rate that can be used on that link). We define the cost of link $l_{ij}$ as $c_{ij} = S_i/r_{ij}$, where $S_i$ is the number of flows that go through node $i$. We then select the minimum-cost path to each destination, that can be computed by a shortest path algorithm, such as Dijkstra's or Bellman-Ford's algorithm.

## 10.6 FRSA Implementation as an Extension of the OLSR Scheme

In this section, we briefly summarize the main features of the OLSR protocol, and then describe how we extend the OLSR scheme to apply our FRSA, leading to a relay-quality aware routing. Finally, we show some performance results obtained via simulation.

### 10.6.1 Background on OLSR

OLSR is a proactive link state protocol, which involves regular exchange of topology information among the network nodes. It employs designated nodes called *Multi Point Relays (MPRs)* to facilitate controlled flooding of topology information. MPRs are also the sole constituent nodes in the route between any source-destination pair in the network.

**HELLO Message Broadcast and Processing.** Every OLSR node periodically broadcasts heartbeat HELLO messages, with information about its neighbors and the corresponding link states. A link state can be symmetric, asymmetric or MPR. An MPR link state with a neighbor indicates that the neighbor has been selected by this node as an MPR; MPR links are symmetric. The HELLO messages are broadcast to all one-hop neighbors, but are not relayed to nodes which are further away. A *Neighbor Table* at each node stores the information about the one-hop neighbors. Upon receiving a HELLO message, a node creates or updates the neighbor entry corresponding to the node which sent the message.

**Multipoint Relays.** Based on the information obtained from the HELLO message, each node in the network selects a set of nodes amongst its symmetrically-linked neighbors, that help in controlled flooding of broadcast messages. This set of nodes is called the Multipoint Relay set of the node. The neighbors of the node which are not in its MPR set, receive and process broadcast messages from the node, but do not retransmit them. The MPR set is selected such that it covers all the nodes that are two hops away.

**Topology Control Message Broadcast and Processing.** The *Topology Control (TC)* messages are broadcast by a node in the network to declare its *Multipoint Relay*

*Selector Set (MPRSS)*. The MPRSS of a node $x$ includes the nodes that have selected $x$ as an MPR. A node obtains MPRSS information from periodic HELLO messages received from its neighbors. Each node in the network maintains a *Topology Table*, in which it records the information about network topology as obtained from the TC messages. An entry in the topology table contains the destination address, $T_{dest}$, and the address of the last hop to the destination, $T_{last}$. Each such entry means that node $T_{dest}$ has selected node $T_{last}$ as an MPR and that node $T_{last}$ has announced this information through a TC message.

**Routing Table Calculation.** The routing table is evaluated based on the connectivity information in the neighbor table and topology table. Shortest path algorithm is employed for route calculation. Each resulting route entry consists of the destination entry, the next node from the sender, and number of hops to the destination.

### 10.6.2 Relay Quality-Aware Routing

In principle, in order to perform a relay quality-aware routing, a node would need a complete knowledge of the network, in terms of nodes, connected pairs and relative rates. Since acquiring such knowledge is unrealistic, we will describe a scalable approach below.

The measurement of the data transmission rate requires cross-layer interaction between the routing protocol and the MAC layer, which is common to many recent proposals. However, the MAC layer evaluation of the maximum rate between nodes raises a further issue. Since broadcast transmissions are performed by 802.11 at 1 Mb/s, determining the actual maximum achievable rate between node pairs requires unicast transmissions. Sending unicast data to every node within transmission range would introduce complexity, overhead and unreliability (e.g., latency in detecting topology changes). Our idea, instead, exploits the signal to noise plus interference ratio (SINR) to estimate the available data transmission rate. This information can be obtained at the MAC layer at each packet reception, no matter the data transmission rate employed by the sender. Thus, a single broadcast transmission allows all neighbors to estimate the quality of the channel from the originating node, at the same time. Note that this approach is especially fit to proactive routing such as OLSR: indeed, the topology control messages each node is required to send at short, regular time intervals allow a comprehensive and frequently updated SINR estimation for each neighbor. Recorded SINR information is handed over to the routing protocol, which smooths it through an Exponential Moving Average filter to avoid short–term effects, and extrapolates the corresponding data rate by looking at SINR thresholds, which separate the working intervals of the different channel coding techniques [9].

The rate resulting from the previous computation refers to the link from the neighbor node to the node which receives the data, i.e., the *reverse* link. Since link symmetry is not guaranteed in a wireless environment, this value could not correspond to the rate achievable from the current node to its neighbor. Thus, the information about the reverse link quality must be communicated back to the neighbor that generated the transmission.

Once a node receives the reverse link rate information from a neighbor (corresponding to its *forward* rate to the neighbor), it can add this information to its knowledge base, and notify it along with other topology information in the control messages that are necessary to the proactive routing functioning.

Referring to the notation introduced in Section 10.4, we implemented the cost function as $c_{ij} = K + 1/(r_{ij} \cdot w_i)$, where $K$ is a constant, $r_{ij}$ is the data transmission rate from node $i$ to node $j$, and $w_i$ is the *relay willingness* of node $i$, i.e., a measure of the willingness of the node to act as a relay for the data coming from an additional node. The additive constant $K$ weighs each extra hop by an empirically determined value of 0.25, so as to fulfill the guidelines listed in Section 10.4. The relay willingness is locally set by each node considering factors such as the locally generated traffic, the already relayed traffic, the level of mobility, the selfishness of the user. As an example, a highly mobile node should not act as a relay, since its link will often break. The relay willingness information must be advertised by nodes, and, again, this can be easily done through topology control messages.

Next, we describe how OLSR can be easily extended so that each node performs a relay quality-aware routing limited to its two-hop neighborhood, and not applied to the whole network. This means that, for nodes that are farther than two hops away, a node uses the standard OLSR routing table computation. This choice is justified by the fact that, once the forwarded data exits the two-hop neighborhood of the originator, the node that relayed the data last can apply the relay quality-aware routing over the next two hops, and so on.

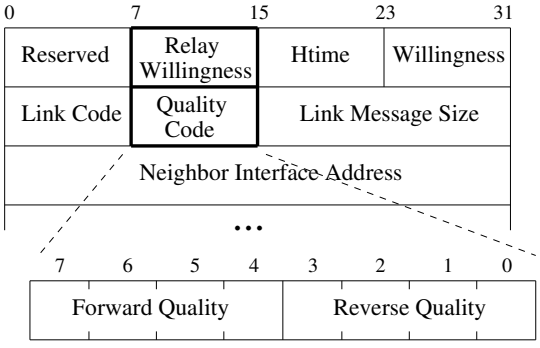



**Fig. 10.23.** Modified HELLO message format.

Our design for a relay quality-aware routing extension to OLSR does not require any change in the structure of control messages. However, *reserved* fields of the HELLO message format are used to exchange link quality and relay willingness information, as described below and shown in Fig. 10.23.

- An 8-bit *Relay Willingness* field, advertising the willingness of the node to act as a relay for data flows. This information is used to compute the link cost, as described before.

- An 8-bit *Quality Code*, advertising the forward and reverse link quality to and from the neighbor nodes, characterized by the given Link Code, listed below. The 4-bit forward link quality is used to compute the cost of the link, while the 4-bit reverse link quality lets each neighbor know its forward quality to the node originating the message, as previously discussed. Four bits allow a wide range of choices: in the simplest implementation, five codes are allowed, corresponding to the four data rates provided by 802.11b.

The following changes to the record structure defined by the standard are also needed.

- Two fields, *L_fwd_quality* and *L_rev_quality*, are added to the Link Tuple format, storing forward and reverse link quality. The reverse quality value is computed from the SINR observed for the current link, while the forward link quality is obtained from HELLO messages.
- A *N_fwd_quality* field is added to the 2-hop neighbor set, storing the quality of the link from the node neighbor to the two-hop node neighbor, and is used to compute the cost of the link.

Clearly, this scheme, limiting the relay quality-aware routing to two times the transmission range, could generate sub-optimal results. An extension of the relay quality-aware routing to the whole network is also possible, by operating an MPR selection based on measured and advertised link quality scores, similar to [7], and then introducing link quality information on OLSR TC messages.

### 10.6.3 Simulation Results

We tested FRSA on three topologies, shown in Fig. 10.24, 10.25 and 10.26, that we dubbed, respectively, the *parking lot*, the *fork*, and the *fan* topology.

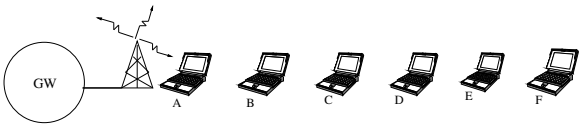



**Fig. 10.24.** Parking lot topology.

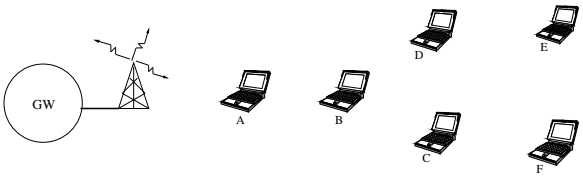



**Fig. 10.25.** Fork topology.

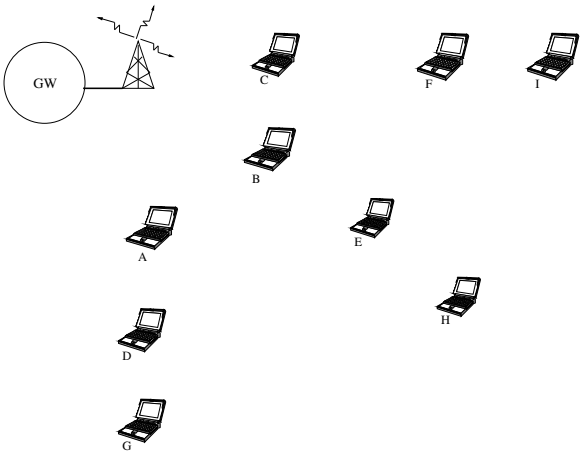



**Fig. 10.26.** Fan topology.

Where not specified differently, it is assumed that propagation conditions are such that nodes that are adjacent to each other in the above figures can achieve *on average* a reciprocal transmission rate of 11 Mb/s; the rate between two wireless nodes falls to 5.5 Mb/s if there is one intermediate node among them, and to 2 Mb/s and 1 Mb/s in case of two or three intermediate nodes, respectively. Nodes farther than that cannot decode each other's transmissions, but they still sense each other, thus all of the presented scenarios identify a *clique* when the relative interference graph is taken into account.

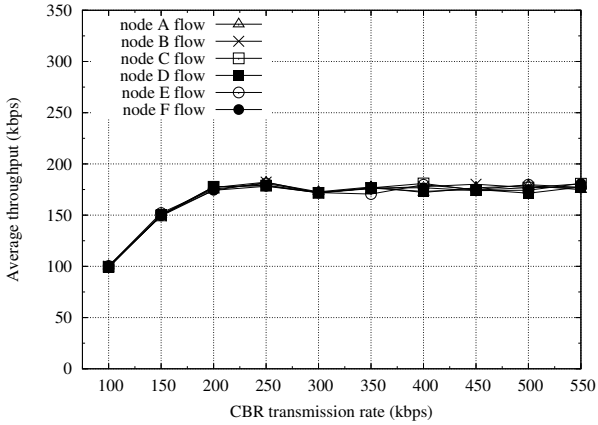



**Fig. 10.27.** Single CBR flow throughput with standard OLSR (parking lot topology).

Therefore, as an example, node B in the fork topology can communicate at 11 Mb/s with both C and D, and at 5.5 Mb/s with both E and F. As already pointed

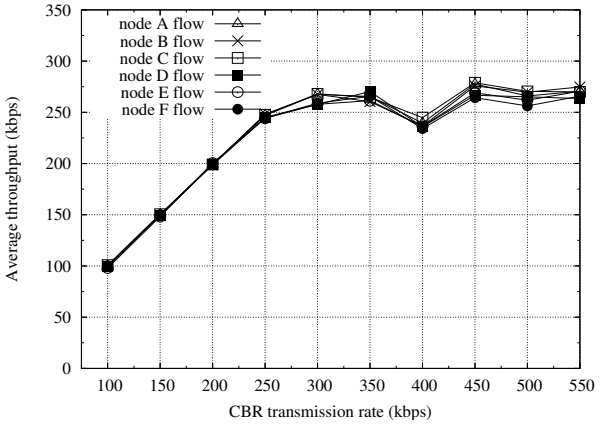



**Fig. 10.28.** Single CBR flow throughput with FRSA (parking lot topology).

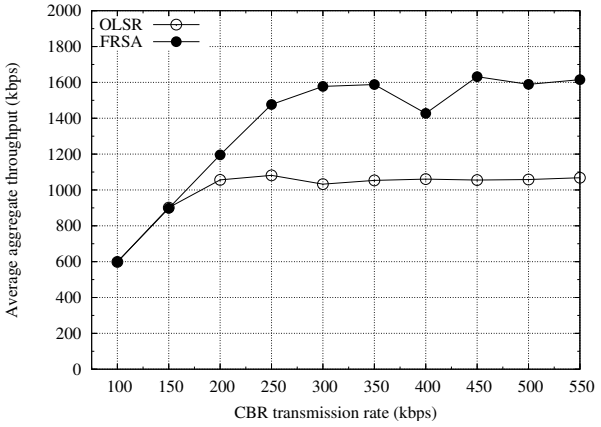



**Fig. 10.29.** Aggregate throughput comparison between standard OLSR and FRSA (parking lot topology).

out, these are average values, since the coupled simulation of ARF techniques at nodes and propagation-dependent channel errors force the nodes to vary their rates over time. All traffic is supposed to be of CBR over UDP nature and flowing in the downlink direction, i.e., from the mesh gateway (GW) to the mesh nodes.

The first set of results refers to the parking lot topology and it shows the average throughput achieved by every node as a function of the source transmission rate. Simulations are performed using standard OLSR (Fig. 10.27) and FRSA (Fig. 10.28), and they show that a significant throughput increase (over $50\%$) can be achieved by FRSA, while maintaining fair access. Fig. 10.29 provides additional proof of our claims by reporting the aggregate throughput.

The fork topology illustrates a special case: due to their distance from the AP, no node can achieve 11 Mb/s on its direct link from the AP, the node that enjoys the highest rate from the AP being node A, with a 5.5 Mb/s transmission rate. Standard OLSR (Fig. 10.30) has nodes C, D communicate directly with the AP at 1 Mb/s, while it has nodes E, F use D as their relay, achieving a combined 11 Mb/s + 2 Mb/s rate. This is clearly a suboptimal solution, since it lowers all transmission rates, hence decreasing the overall performance due to the anomaly. FRSA, on the contrary, has C, D use A as their relay, and E, F use B as their relay; in either case, the rates are a combined 5.5 Mb/s + 5.5 Mb/s. Since no node has the upper hand in terms of transmission rates, the balance is tipped in favour of those nodes, namely A and B, who receive their traffic through fewer hops (Fig. 10.31). However, the overall throughput of the system again experiences a $50\%$ rise, as shown in Fig. 10.32.

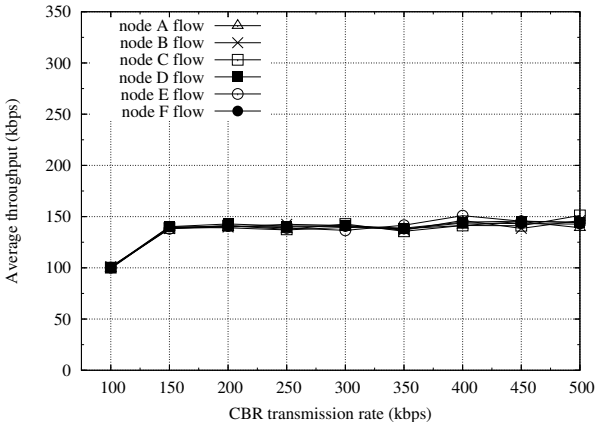



**Fig. 10.30.** Single CBR flow throughput with standard OLSR (fork topology).

The fan topology defines three concentric areas around the AP. Nodes on the same branch are referred to as *relative* nodes. The first rim, including nodes A, B and C, can achieve a 5.5 Mb/s direct rate to the AP; the second rim, including nodes D, E and F, has 1 Mb/s direct rate to the AP, or a 5.5 Mb/s rate toward their relative inner rim node; finally, no direct link can be established between nodes on the third, outer rim and the AP. These nodes can instead establish a 1 Mb/s link toward their relative first rim node, or a 5.5 Mb/s link toward their relative intermediate rim node. Note that these rates only refer to relative nodes, i.e., two nodes on separate branches can experience different rates with respect to those mentioned above, depending on their distance. In this case, standard OLSR can route traffic to nodes on the third rim through any of the six nodes in the inner rims, and, since it just considers the hop count as the cost metric, all the six inner rim nodes are considered as equivalent. This can bring standard OLSR to select paths which are trivially sub-optimal. As an example, in one of our simulations, standard OLSR initially routed traffic directed to node G through node B, while node A would have been a better relay in any case,

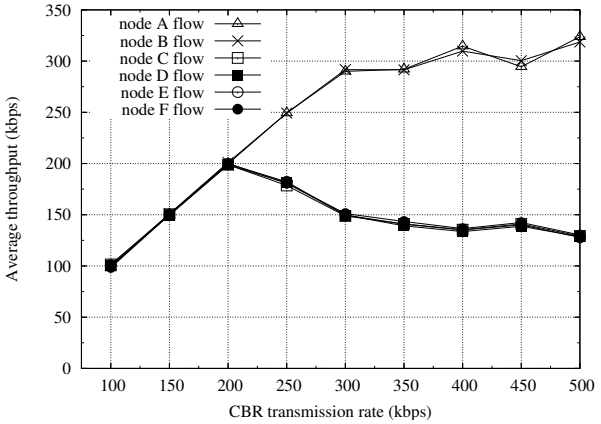



**Fig. 10.31.** Single CBR flow throughput with FRSA (fork topology).

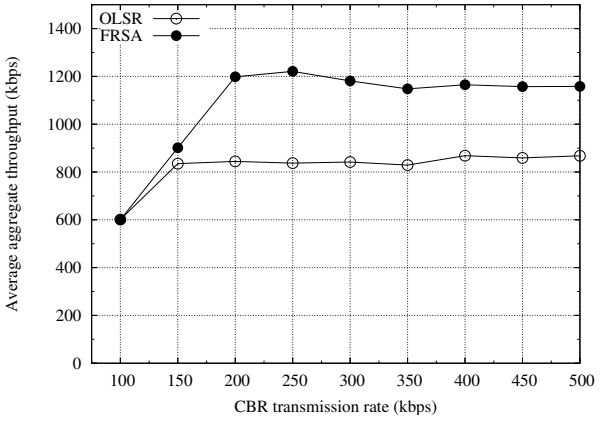



**Fig. 10.32.** Aggregate throughput comparison between standard OLSR and FRSA (fork topology).

its distance from G being clearly lower than B's. Moreover, standard OLSR's lack of preference in the relays toward the outer rim nodes brings the network to very high route instability, when traffic saturation conditions are reached. As a matter of fact, the losses deriving from high contention and full buffers involve OLSR routing messages and lead to very frequent path changes. We noticed that, under such conditions, it is not possible to identify clear routing paths anymore, as the nodes continuously modify their routing tables, struggling for reliable links. Obviously, it would be desirable to avoid such a behavior, as it only adds overhead and definitely does not bring any advantage to the network.

When FRSA is employed, we first of all observe an optimal route selection, as the fast 5.5 Mb/s links are fully exploited at the expenses of the 1 Mb/s links, which

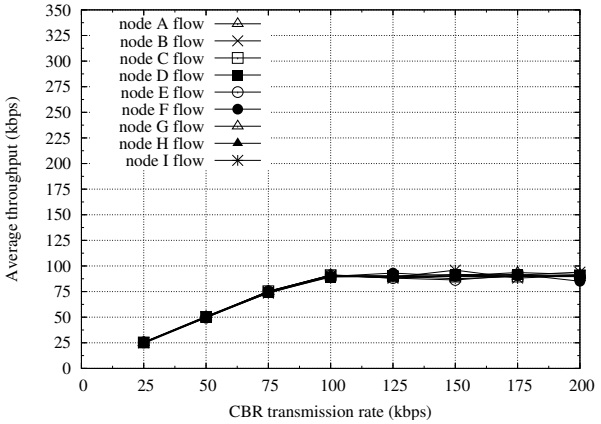



**Fig. 10.33.** Single CBR flow throughput with standard OLSR (fan topology).

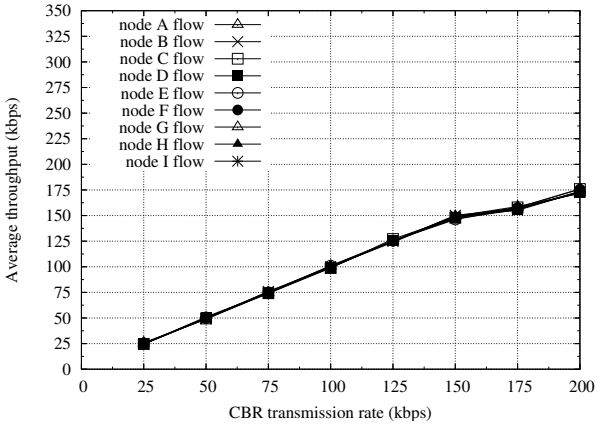



**Fig. 10.34.** Single CBR flow throughput with FRSA (fan topology).

are avoided instead. This means that each node uses its relative inner rim neighbor to route the traffic, with the middle rim nodes reached through a two-hop path and the outer rim nodes reached via a three-hop route. Secondly, simulation results show that this distribution of independent data flows along the different branches of the fan is not affected by the traffic load. The routes are rarely changed, even when the system reaches saturation, as FRSA manages to identify that the most profitable paths do not change as the uniform traffic load increases. The beneficial effect of FRSA is evident when comparing Fig. 10.33 and Fig. 10.34, showing the per-flow throughput obtained in the two cases: when FRSA is used, fairness is maintained but the network performances are noticeably improved, leading to an aggregate throughput, depicted in Fig. 10.35, nearly doubled at high loads.

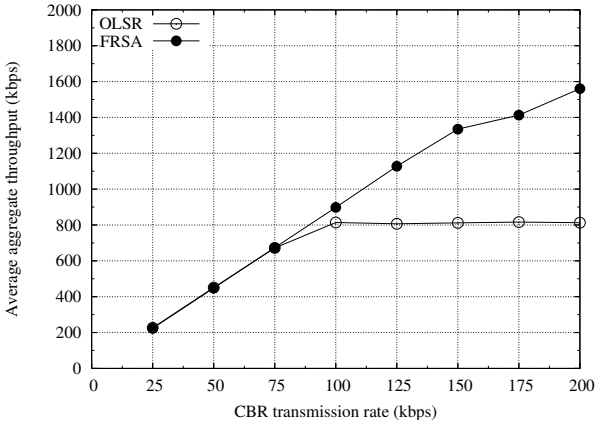



**Fig. 10.35.** Aggregate throughput comparison between standard OLSR and FRSA (fan topology).

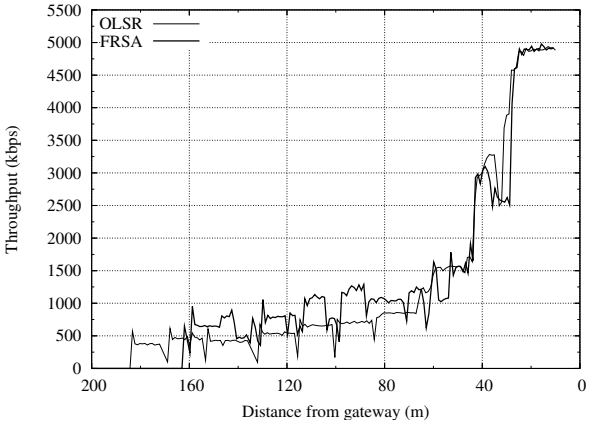



**Fig. 10.36.** Temporal diagram of aggregate throughput of standard OLSR and FRSA (parking lot topology with mobile node).

In the fourth scenario, we tested a mobile environment where a node is moving on a straight line toward the AP, along the parking lot topology. The mobile node starts outside radio range of node F and inches toward the AP, receiving traffic as soon as it can establish a link with any node in the neighborhood. All other nodes are supposed to have no traffic of their own, but they relay traffic to the mobile node, if needed. Fig. 10.36 shows the achievable throughput over time as the node cruises along the topology toward the AP, again showing the higher throughput achieved by FRSA. The histograms in Fig. 10.37 and Fig. 10.38 show the distribution of relay choices with standard OLSR and FRSA, respectively; each bar reports the fraction of packets routed through the mesh node at each achievable rate.

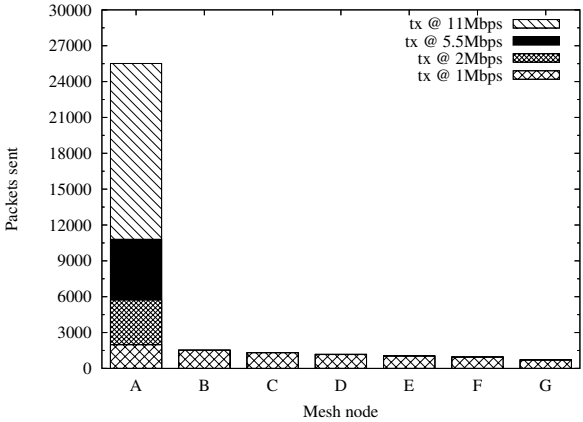



**Fig. 10.37.** Distribution of relay choices with standard OLSR (parking lot topology with mobile node).

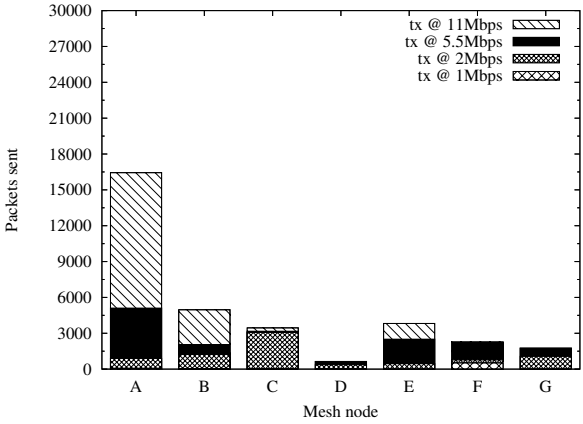



**Fig. 10.38.** Distribution of relay choices with FRSA (mobile topology).

When running standard OLSR, the AP privileges routing downlink traffic through the closest neighbor that has a direct link toward the mobile node. As a result, the algorithm tends to use the lowest rates, as they guarantee the greatest reach. On the other hand, FRSA alternately picks all nodes of the topology exploiting the best available links. The data transmission is only routed directly to the mobile node when the link exhibits a sufficiently high quality.

## Conclusion

This chapter has provided some perspectives on issues related to mesh networks. First, the anomaly effect in 802.11-based networks has been outlined and the use

of multihop transmission through traffic relays has been introduced as a countermeasure. Then, two relay strategies have been described, which aims at exploiting MAC/routing interactions in an 802.11-based, multi-rate WLAN. The presented techniques give incentives to relay nodes so that high-throughput multihop communications can take place. Finally, a Fair Relay Selection Algorithm (FRSA) has been proposed and implemented as an extension of the OLSR routing protocol. Following the guidelines laid out by empirical observations, FRSA aims at an efficient selection of relay nodes in an automated fashion. An extensive set of simulation results have showed that higher throughput and fairer channel access among all mesh nodes are achievable when FRSA is used instead of OLSR in both static and mobile mesh scenarios.

## References

1. I. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *Computer Networks*, vol. 47, no. 4, pp. 445-487, Mar. 2005.
2. D. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multihop wireless routing," in *Proc. IEEE/ACM MobiCom 2003,* San Diego, CA, pp. 134-146, 2003.
3. C.-F. Chiasserini and M. Meo, "An innovative routing scheme for 802.11-based multihop networks," in *Proc. IEEE VTC Fall 2004,* Los Angeles, CA, pp. 2804-2807, Sept. 2004.
4. M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. IEEE Infocom'03,*, San Francisco, CA, pp. 836-843, 2003.
5. IEEE 802.11 WG Draft Supplement to Standard Part II: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: MAC Enhancements for Quality of Service, IEEE 802.11e Draft 11.0, 2004.
6. C. Casetti, C. F. Chiasserini, and M. Fiore, "Relay quality awareness in mesh networks routing," in *Proc. Tyrrhenian International Workshop on Digital Communications (TWD'07)*, Ischia Island, Italy, Sept. 2007.
7. H. Badis, A. Munaretto, K. Al Aghal, and G. Pujolle, "Optimal path selection in a link state QoS routing protocol," in *Proc. IEEE VTC Spring 2004,* Milan, Italy, pp. 2570-2574, May 2004.
8. B. Awerbuch, D. Holmer, and H. Rubens, "High throughput route selection in multi-rate ad hoc wireless networks," *Tech. Rep.,* Johns Hopkins University, 2004. http://www.cs.jhu.edu/archipelago/
9. Y. Seok, J. Park, and Y. Choi, "Multi-rate aware routing protocol for mobile ad hoc networks," in *Proc. IEEE VTC 2003-Spring,* Jeju, Korea, pp. 22-25, Apr. 2003.
10. R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multihop wireless mesh networks," in *Proc. IEEE/ACM MobiCom 2004,* Philadelphia, PA, pp. 114-128, 2004.
11. H. Zhu and G. Cao, "On improving the performance of IEEE 802.11 with relay-enabled PCF," *Mobile Networks and Applications,* vol. 9, no. 4, pp. 423-434, 2004.
12. Y.-C. Tse, W. Chu, L.-W. Chen, and C.-M. Yu, "Route throughput analysis for mobile multi-rate wireless ad hoc networks," in *Proc. IEEE 1st International Conference on Broadband Networks (BROADNETS'04)*, San Jose, CA, pp. 469-475, 2004.
13. Iperf [Online] http://dast.nlanr.net/projects/Iperf/
14. R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling,* Wiley-Interscience, New York, NY, 1991.

# 11

# Multiple Antenna Techniques for Wireless Mesh Networks

A. Gkelias and K. K. Leung

Imperial College, London, UK
{a.gkelias,kin.leung}@imperial.ac.uk

## 11.1 Introduction

Wireless mesh networks (WMNs) is a relatively new and promising key technology for next generation wireless networking that have recently attracted both the academic and industrial interest. Mesh networks are expected gradually to partially substitute the wired network infrastructure functionality by being able to provide a cheap, quick and efficient solution for wireless data networking in urban, suburban and even rural environments. Their popularity comes from the fact that they are self-organized, self-configurable and easily adaptable to different traffic requirements and network changes. Mesh networks are composed of static wireless nodes that have ample energy supply. Each node operates not only as an conventional access point (AP)/gateway to the internet but also as a wireless router (Fig. 11.1) able to relay packets from other nodes without direct access to their destinations [1] [2]. The destination can be an internet gateway or a mobile user served by another AP in the same mesh network. Moreover, some nodes may only have the backhauling functionality, meaning that they do not serve any mobile user directly but their purpose is to forward other APs' packets.

For wide area access, the access points (or base stations) are typically located at high towers or at the rooftop of buildings. However, as the capacity demands increase, the AP is moving closer to the user and it could be placed at below-the-rooftop heights. In this way it can provide better signal reception and higher spatial frequency reuse factor. For a cellular network point of view this means that the cell size has to shrink in order to satisfy the increased capacity demands. Therefore, the whole network topology in terms of base station location has to be revised and additional base stations have to be installed by the network operator. In case where fiber is not readily available, the cost of backhauling by using e.g. E3 lines may be prohibitive for the operator. Wireless mesh networks can be proven an appealing alternative backhauling solution in this case since they do not require wired connection, they are easy to deploy fast and without extensive network planning requirements since they are self configurable and adaptable to network changes and demands. For instance, Multi-Element Multihop Backhaul Reconfigurable Antenna Network

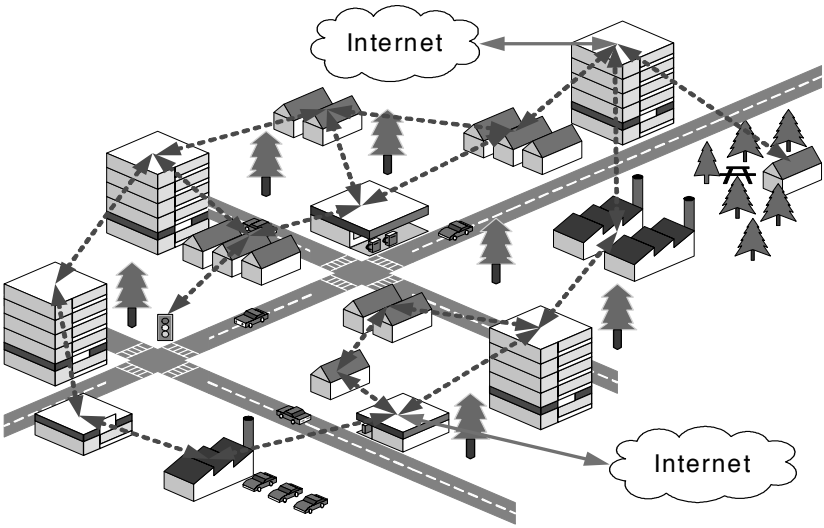



**Fig. 11.1.** Conceptual illustration of a multi-hop wireless mesh network.

(MEMBRANE) [3] is an IST-funded project that aims to bring an efficient wireless backhaul design as an alternative technology to serve wireless broadband networks in cases where a wired backhaul would be more costly to access and/or would take longer to deploy.

Moreover, wireless mesh network can enhance the presence of broadband in rural and remote zones, thus helping combat the "digital divide" between these areas and the big urban centers, caused mainly by the inadequacy (or even absence) of wired network infrastructure. It is not only doubtful that the required investment for bringing cable and/or fiber will ever pay for itself in such remote zones and communities but, more importantly, such an undertaking will probably take several years. Wireless mesh networks can provide a quick and economically affordable solution in this case.

*Technology Enablers*: Mesh networks must meet a number of technical requirements. First of all, they must meet the high capacity needs of the access nodes which have to forward the accumulated traffic of their underling users. Furthermore, they have to cope with the delay and other strict quality-of-service (QoS) requirements of the end user applications. Finally they must provide a large enough effective communication range to ensure that no APs (or groups of APs) are isolated from the Internet gateways. In order to satisfy the above requirements, a range of novel techniques has to be exploited. Such technology enablers include but not limited to multi-hopping, various multiple antennas techniques and novel medium access control (MAC) and routing algorithms.

*Multi-hopping*, i.e. the use of multiple relays (or forwarding nodes) between the end user and the Internet gateway, is primarily motivated by the low power and the low heights of the access and relay nodes. Clearly, in low power transmissions, multi-hopping helps increase the range. Moreover, since low height access points are likely to be surrounded by several obstacles (buildings, cars, etc.), their line-of-site (LOS) will be typically obstructed, affecting in this way the one-hop node connectivity to a gateway.

*Multiple antenna techniques* (also known as smart antennas) constitute another enabling technology that is highly beneficial to the wireless mesh network architectures. These techniques include fixed beam antennas, adaptive antennas and multiple-input multiple-output (MIMO) coding [4] [5]. Depending on the used technique, multiple antennas can provide power, diversity and multiplexing gain and therefore increase the transmission range, reduce the transmitting power, mitigate interference, increase channel reliability and increase data throughout. Each of these techniques is more appropriate in different types of propagation scenarios; for example beam-forming is well suited to cases with narrow angle spread, such as in high towers; whereas MIMO is more appropriate in cases of rich electromagnetic scattering, such as in low-height links without strong LOS component.

Given the different types of propagation environments that are expected in wireless mesh networks, smart antenna techniques are expected to boost throughput performance and reduce interference and delay, thus improving overall end-to-end performance. However, multiple antenna techniques have been extensively analyzed only for single-link communications. The combination of multi-hopping with multiple smart antennas in a wireless mesh network environment is a field that has not received much research attention. This combination is expected to boost network capacity and achieve the target QoS.

Novel *medium access control and routing algorithms* that are able to exploit the benefits of multiple antenna usage on the wireless mesh access points is another enabling technology of paramount importance. Employment of smart antennas techniques without thorough understanding and consideration of their interaction with layer two and three algorithms can be proven counter productive for the mesh network functionality. Deafness, hidden and exposed terminals and multi-stream interference are some of the problems that have to be addressed by novel MAC and routing schemes since they can highly affect not only the individual links but also the overall network performance.

The main aim of this chapter is to give an insight into both the improvements and various challenges generated by the deployment of multiple antennas in wireless mesh networks and how these can be addressed by layer two and three algorithms. In the rest of this chapter the wireless mesh network and channel characteristics are discussed in Section 11.2. A thorough analysis of the various smart antenna techniques and their main advantages and disadvantages follows in Section 11.3. In Section 11.4, the challenges that smart antenna techniques will impose to medium access control and network layers are introduced together with some possible solutions. Finally, Sections 11.5 and 11.6 discus several scheduling and routing schemes, respectively, with smart antenna considerations.

## 11.2 Channel Characteristics

The wireless channel in a mesh network is expected to be highly dynamic. The dynamic nature of the channel comes both from environmental changes/ movements and from the interference fluctuations from network transmissions. In this chapter, we consider only the access point (AP) to AP communication (rather than the AP to user communication); therefore, we do not expect to have any explicit mobility in terms of transmitter or receiver movement. Nevertheless, the wireless mesh networks are expected to be deployed in urban or suburban environments where the surrounding objects (cars, trains, and people) may be in constant move. Also nodes failure can affect the network connectivity by causing to higher layers similar effects as node mobility.

### 11.2.1 Propagation Scenarios

Unlike legacy cellular networks, where base stations are exclusively mounted on high towers or at the rooftops of tall buildings after extensive network planning, aiming for high LOS coverage, wireless mesh APs can be less neatly deployed. In order to reduce the deployment cost and being closer to the users, the APs and relay nodes are mostly placed at low-to-moderate heights, in order of 3-10m (for example mounted on electrical and telephone poles, traffic lights, building sidewalls and rooftops) where direct LOS is difficult to be guaranteed. Depending on the relative position of the AP we can have different communication scenarios that highly affect the channel propagation statistics [3]:

**Rooftop to Rooftop:** In this scenario both ends of a link are placed above the rooftop level. Pure LoS conditions are met as far as the first Fresnel zone is clear.

**Below-rooftop to Below-rooftop:** This refers to the case where both nodes are deployed below the level of the surrounding buildings and it covers both LoS and NLoS outdoor propagation conditions.

**Rooftop to Below-rooftop:** In this scenario the one end of the link is located above the rooftop level while the other is below that. This case possesses strong similarities with the traditional cellular case. The major difference is the Doppler spectrum shape and that the APs are placed at moderate height. Although LoS conditions are possible, the NLoS case is more probable.

Each scenario has an important impact in various channel properties, such as path-loss, angle and delay spread and highly affect the optimum multiple antenna technique that should be used. For instance, at large scattering angles, MIMO performs better than adaptive beamforming techniques. However, at low-moderate scattering angles adaptive or even some simple switched beam techniques can be more beneficial than MIMO.

### 11.2.2 Power Constraints

Since the mesh network transceivers are usually located on the top or side-walls of buildings or special contracted poles that have easy access to power through wires,

power supply and consumption is not a crucial issue for mesh network unlike sensor and mobile ad hoc networks. Moreover, since they are located close to the users, and for health and safety reasons, the wireless mesh transceivers are expected to operate at relatively low powers (on the order of at most a few Watts). Last but not the least, total effective radiation power limitations may apply in different countries for directional transmissions in unlicensed bands (where a mesh network can operate) as it will be discussed in a latter section.

### 11.2.3 Interference Characteristics

Inevitably, due to the spatial channel reuse in wireless mesh networks a given node will suffer (co-channel) interference from other nodes making the wireless communication more interference limited rather than noise limited. In multiple-antenna systems it is not only the signal-to-interference-plus-noise ratio (SINR) that affects the network capacity but also the distribution of the interference power. In a mesh network, the interference is not spatially white but it will rather emanate unequally from different directions (spatial color). Recent results in [6] and [7] have shown that MIMO systems for a given SINR perform more efficiently the more spatially colored the interference is (i.e., it is better to have few and high-power interference components rather that many low-power ones).

Moreover, in urban highways, big buildings in both sides act as natural obstacles that can waveguide the signal and significantly reduce the interference from/towards adjacent streets. The degree that buildings and natural obstacles can affect the signal propagation and reception highly depends on the carrier frequency of the signal. These interference characteristics have to be taken into account for the optimum design of wireless mesh network algorithms.

## 11.3 Smart Antenna Techniques

The use of intelligent antennas in ad hoc networks has recently attracted a great amount of attention as a means to optimize power transmission/reception ( [8] and references therein). Two basic types of intelligent antennas are considered in this context: directional antennas (fixed beams) and adaptive antenna arrays (also known as smart antennas). A directional antenna generates multiple pre-defined fixed beam patterns and applies one at a time towards the direction of interest. It is the simplest technique, essentially providing sectorisation with the capability of illuminating the selected sector according to, for instance, an SINR-related metric. An adaptive antenna array can formulate the beam structure based on a certain optimization criterion, such as maximizing the array gain towards the signal of interest and suppressing interfering signals. MIMO techniques can be seen as an extension of adaptive antenna arrays. They require multiple antennas at both end of the link and are capable to provide spatial multiplexing or diversity gain. In this section, we a) summarize the different multiple antenna techniques, b) give an insight into the tradeoffs of the various performance gains they can achieve and c) discuss the cases (e.g. channel conditions, network requirements) that their usage would be more appropriate.

### 11.3.1 Directional Antenna Techniques

**Switched-Beam Antennas:** Switched beam is the simplest technique. A predetermined antenna array pattern or separate directive antennas are used to generate a limited number of beams that point to desired directions. These beams can be used either for transmission or reception and each time a beam-switching algorithm determines which particular beam will be used to maintain the highest quality signal. The predefined beams can be switched in a mechanical or electronic way. This ability to concentrate power in a certain direction provides a directive gain (also called power gain or array gain) that can be used for extending range or reducing power. This type of antenna is easy to be implemented (e.g. using multiple antenna elements, each pointing to different direction, where direction is chosen by choosing the element), but it gives a limited improvement.

**Steered-Beam Antennas (or Dynamically Phased Arrays):** Steered beam antennas have also predefined patterns but they can be pointed to any of a near continuous set of directions. This can be achieved by phase shifting and combining the signals emitted from each element of an antenna array. Direction of arrival (DoA) techniques can be used to continuously track the direction of the receiving signal and steer the beam accordingly [4]. This helps to avoid the performance degradation occurred in switched beam antennas due to "scalloping loss" [9] (the degradation due to scalloping loss is more significant in mobile environments rather than in static mesh networks). Although any arrangement of antennas can be used, the most typical would be linear, circular, and planar arrays [12].

While directional antenna techniques can provide sufficient gain in terms of SINR in presence of strong line-of-sight component and no interference, their performance deteriorates significantly in multi-path environments where the desirable signal can arrive from multiple directions.

### 11.3.2 Adaptive Antenna Techniques

Adaptive antenna arrays, or smart antennas [10], use a combination of an array of multiple antennas and appropriate signal processing to produce desirable antenna patterns. Such patterns have high gain in the direction of desired signals and nulls in the direction of undesired signals.

**Adaptive Antenna Arrays:** Adaptive antenna array at the receiver can provide power gain (array gain) by coherent combining the received signal copies from all antenna elements. The effective total received signal power increases linearly with the number of antenna elements. Furthermore, its radiation patterns can be adjusted to null out the interference from other directions [11]. In order to achieve this, intelligent digital signal processing (DSP) algorithms should be used [15] to estimate the DoA (several direction of arrival estimation techniques such as ESPRIT [13] and MUSIC [14] can be used) of all the impinging signals, both signal of interest and interfering signals. Interference suppression is obtained by steering beam pattern nulls in the direction of the interfering signals while maintaining the main lobe in the direction of the desired signal. For an antenna array with $N$ elements and for

$M$ interfering nodes ($M < N$), $M$ nulls can be formed to eliminate the receiving power of $N$ separate interference while the remaining $N - M$ antennas can be used to beamform towards the direction of the desired signal. Finally, at the transmitter side the adaptive array can form a narrow beam towards the direction of the desired receiver while optimally suppresses the interference towards any other possible adjacent receivers.

**MIMO Techniques:** MIMO systems can be viewed as an extension of the smart antenna techniques described above. The main characteristic of MIMO techniques is their ability to exploit multi-path propagation rather than mitigate it. They take advantage of random channel fading and multi-path delay spread for multiplying transfer rates (multiplexing gain) or improve the transmission quality/reliability (diversity gain) at no cost of extra spectrum (only hardware and complexity are added). In this way it transforms a traditionally pitfall of wireless channel into a benefit of the communication system.

*CSI*: There is a large number of transmission and reception schemes over MIMO channels depending on the channel state information (CSI) available at the transmitter and/or receiver side and on the diversity and/or multiplexing gain that has to be achieved. CSI at the receiver end (CSIR) can be obtained fairly easy by sending a pilot symbol for channel estimation. At the transmitter side, channel state information (CSIT) can be obtained by CSIR transmission from the receiver through a feedback channel. Alternatively, CSIT can be estimated in bi-directional systems without feedback by exploiting the reciprocal properties of the channel. The former method introduces a trade-off between feedback channel bandwidth and CSI accuracy while the latter one cannot be applied in communication systems, such as frequency-division duplex (FDD) systems, where the reciprocal property does not hold.

*Spatial Diversity Gain:* Antenna diversity (or spatial diversity) can be achieved by placing multiple elements at the receiver and/or the transmitter. These antenna elements need to be placed sufficiently far apart such that the received signal replicas from different antenna elements fade more or less independently. In this case, there is a high probability that at least one or more of these signal components will not be in a deep fade that reduces the variance of the SNR. The required antenna separation depends on the local scattering environment as well as on the carrier frequency. The spatial diversity gain depends on the diversity order, which in turn depends on the degree of which the multi-path fading on the different antenna elements is uncorrelated. For $M$ transmit and $N$ receive antennas a maximum diversity gain of $MN$ can be achieved.

Transmit diversity can be achieved via space-time-coding (STC) schemes. The simplest STC scheme is the Alamouti scheme [16], designed for two transmit and two receive antennas without any feedback from the receiver and is one of the most popular techniques proposed in several third-generation cellular standards for transmit diversity. Space-time block coding (STBC), introduced in [17], generalizes the Alamouti scheme to an arbitrary number of transmit antennas and is able to achieve the full diversity promised by the transmit and receive antennas. At the receiver end linear processing maximum likelihood (ML) decoding is used to decouple the signals transmitted from different antennas and perfect CSIR is assumed.

*Spatial Multiplexing Gain:* In the presence of multi-path or rich scattering, a MIMO system can provide spatial multiplexing gain. This can be achieved by simply sending independent data streams over each of the transmit antenna elements. This technique is known as Vertical Bell Labs Space-Time Architecture (V-BLAST) [18] and it is the first spatial multiplexing technique implemented in real-time in a laboratory. For $M$ transmit and $N$ receive antennas, and under fast fading channel conditions, $\min(M, N)$ independent data stream can be transmitted (often referred to as degrees of freedom (DOFs)). In slow fading channel though, the V-BLAST architecture is strictly suboptimal. Another architecture, i.e., Diagonal Bell Labs Space-Time Architecture (D-BLAST) [19] can provide significant improvements by coding and interleaving the code-words across the antennas. However, the diagonal approach suffers from certain implementation complexities. The receiver must multiplex the signals in order to reconstruct the transmitting symbols. Maximum likelihood (ML) decoding is an optimal solution in the sense that it compares all the possible combinations of the symbols, but its complexity increases with the modulation order. Other popular techniques include zero forcing and minimum mean square error estimation combined with successive interference cancelation (MMSE-SIC) [20]. Diversity and multiplexing gain can be simultaneously obtained for a given multiple-antenna channel, but there is a fundamental tradeoff between how much each coding scheme can get [21].

### 11.3.3 Space-Division Multiple Access (SDMA)

In a wireless network scenario with several nodes communicating to a common receiver, multiple receive antennas also allow the spatial separation of the signals of different transceivers, thus providing a multiple-access gain. This use of multiple antennas is also called space-division multiple access (SDMA). The fact that a MIMO receiver can isolate and decode $\min(M, N)$ independent data streams ($M$ transmit and $N$ receive antennas), can be extended to the case of single receiver employed with $N$ antennas and $M$ independent transmitters with one antenna element. This can be generalized to the case of several transmitters using overall $M$ antenna elements in any possible combination given that the received streams are independent.

Based on the communication scenario (e.g., LOS or NLOS), the nature of the channel (e.g., fast or slow fading, high or low SNR) and the type of gain (e.g., power, range, diversity, multiplexing) that has to be achieved, the multiple antenna transceivers have to choose the optimum technique for communication [22]. However, in wireless mesh networks, the choice of a multiple antenna technique in a single link can highly influence the respective decisions of the adjusted links and affect in this way the overall network performance. In this point, it is the responsibility of the higher layers to assist and coordinate the individual link decisions based on one-hop and end-to-end characteristics and requirements in order to harmonize the network functionality. Therefore, it is of paramount importance to design novel medium access control and routing schemes that are able to apply these techniques to the dynamic wireless mesh networks. The deployment of multiple antenna

techniques gives to this design an additional degree of freedom compared to the traditional omni-directional transceivers and makes it an extremely interesting and challenging task. In the remaining of this chapter we give a description of the several research and implementation challenges that come along with the usage of multiple antennas in wireless mesh networks (Section 11.4) and describe several scheduling (Section 11.5) and routing (Section 11.6) solutions proposed in the literature.

## 11.4 Smart Antenna Challenges and Design Criteria for Mesh Protocols

While antenna arrays can provide numerous advantages to wireless communications as discussed above, their introduction in wireless mesh network communication has to be thoroughly investigated since inappropriate or undesirable interaction with higher layers can lead to overall performance degradation. In the following of this section, we introduce and analyze all these issues and research challenges that have to be taken into account during the design of medium access control and routing protocols that used in conjunction with different antenna array techniques. It will be clear that some of the existing layer two and three traditional protocols cannot be used unmodified since their performance will be much worse compared to the omni-directional transmission.

In order to simplify the representation of different communication scenarios we use the same terminology as in [23] to define the distance between two neighbor nodes:

*Omni-Omni (OO) Neighbors:* Nodes can directionally communicate with each other even if both of them are in omni-directional mode.

*Omni-Directional (OD) Neighbors:* Nodes can directionally communicate with each other even if one of them is in omni-directional mode and the second one in directional mode, pointing the first one.

*Directional-Directional (DD) Neighbors:* Nodes can directionally communicate with each other only if both of the nodes are using directional transmission/reception and pointing each other.

Note here that two OO-nodes can achieve OD and DD communication and an OD pair can perform DD communication but not the other way around.

### 11.4.1 Deafness

Deafness is a common problem that arises due to the use of directional antenna techniques and it occurs when a transmitter fails to communicate to its intended receiver, because the receiver beamforms towards a direction away from the transmitter [24]. Therefore, a receiver can increase its power gain from the direction of its main beam(s), but at the same time it becomes deaf in all the remaining directions. For example, in Fig. 11.2, node A is unable to hear nodes B transmission since node A's main-lobe is shifted to a different direction. Deafness can give rise to other

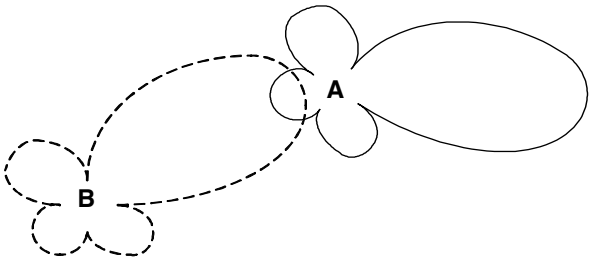



**Fig. 11.2.** Deafness problem.

problems such as the exposed terminal problem and back off fairness. Deafness can be a useful property if proper action is taken into account from higher layers since it can be used for interference mitigation purposes. For instance, a receiver can become deaf to the direction of the impinging interfering signals. In a mesh network, where multiple transmissions take place, new scheduling algorithms have to be defined such that they mitigate the deafness effect for useful signals reception while at the same time take advantage of its interference suppression properties.

### 11.4.2 Hidden/Exposed Terminal Problem Escalation

The hidden/exposed terminal problem is a well known issue in wireless networks [25]. A hidden terminal refers to a terminal which is outside the coverage area of the transmitting node but within the coverage area of the receiving node. A hidden terminal is unable to sense the ongoing transmission, and therefore it may try to transmit and inevitably create interference (or possible packet collision) to the receiving node. Exposed is a terminal that is located within the coverage area of the transmitting node but outside the coverage area of the receiving node. As a result, the exposed terminal will sense the ongoing transmission and will defer its transmission while it should be able to transmit (to another available receiver) since its transmission will not interfere with the ongoing transmission.

Proposed solutions to the hidden/exposed terminal problem are the transmission of a busy tone from the receiver in a separate channel [25] and the exchange of signaling between transmitter and receiver (request-to-send (RTS) and clear-to-send (CTS)) before the actual transmission takes place [26]. The commercial IEEE 802.11 [27] protocols have adopted the RTS/CTS signaling concept from [26] enhanced with the network allocation vector (NAV). While directional transmission can reduce the interference to adjacent nodes at the same time it can augment the hidden and exposed terminal problem if careful consideration of the impact of the beamforming on the medium access control algorithms design is not taken into account [28]. In the following we analyze all these issues related to the usage of directional antennas in carrier sensing multiple access (CSMA) based MAC protocol.

*Initial Channel Sensing*: It is desirable for an idle user to listen to the channel omni-directionally (unless there is a special topology or it knows the location of the

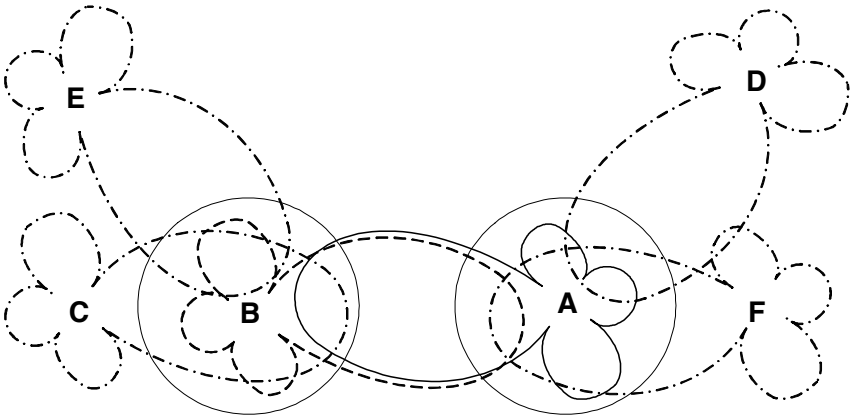



**Fig. 11.3.** A communication example - hidden terminals.

transmitter *a priori*) because a potential transmitter can be located in any possible position around. Moreover, a node has to overhear any other ongoing transmission around its area of coverage. The problem in this case though is that direct communication cannot be established if the transmitter is far enough (DD-neighbors) that cannot be sensed even if directional transmission of RTS takes place. Direct DD-communication could be possible if the receiver is sensing the channel directionally towards the direction of RTS transmission. Therefore, a special mechanism is needed to inform the receiver for the potential transmitter direction. On the other hand, a node that is willing to transmit has to sense the channel in the direction of the receiver in order to avoid hidden terminals in that direction. However, for the time the potential transmitter is in a directional sensing mode it is deaf to any transmission in the direction of its side lobes.

*RTS Transmission*: Since the idle nodes sense the channel omni-directionally a directional RTS packet transmission is needed to initiate communication with an OD-neighbor. Otherwise only the OO-neighbors within its omni-directional coverage area will be able to sense the RTS packet. However, this may generate some hidden terminals in the direction of the transmitter's side lobes.

*Channel Sensing after RTS Transmission (Transmitter)*: After sending a directional RTS packet the potential transmitter senses omni-directionally for CTS reply or any other transmission. This has as an effect the escalation of the exposed terminal since it will sense transmissions from directions that will be eventually suppressed from the side lobes. Furthermore, it will be unable to sense possible transmission from the direction of its receiver that are located outside the omni-directional carrier sensing range but inside the directional range.

For example, in Fig. 11.3, node A sends a directional RTS to node B and switches to omni-mode to sense the channel. Node A will sense the transmission from node D, it will assume that this will overlap with its own transmission and it will defer

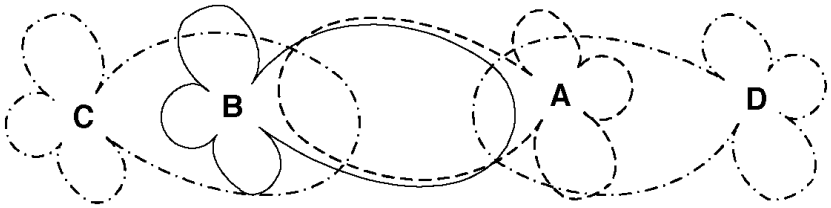



**Fig. 11.4.** Directional CTS communication.

its transmission. However, this transmission will be suppressed by its side lobes and will not affect the A to B communication. On the other hand, transmission from C will not be detected if A is in omni-mode, and will create a collision since A and C are DD-neighbors that point to each other.

Consider now the case that after sending a directional RTS packet the potential transmitter continues to sense at the same direction for CTS packet or any other transmission. This solves the hidden and exposed terminal problem of the previous case but will not avoid hidden terminals from the direction of its side lobes. For example, the node A becomes deaf to the direction of node F. Node F may create interference to node B since they are DD-neighbors that point to each other.

*Channel Sensing before CTS Transmission (Receiver)*: After a node receives an RTS packet it has to determine its DoA and if a packet is destined for itself or any other node. If the packet is intended for the given receiver it can directionally sense the channel towards the estimated DoA for any other ongoing or new transmissions. In this way it reduces the impact of both the hidden terminals problem in the direction of the main beam and the exposed terminals problem from the direction of the side lobes. Collisions can still occur as it is demonstrated by the following example:

Node B sends an RTS to node A (Fig. 11.3) that successfully receives the packet and continues to sense the channel omni-directionally before it replies with a CTS. Node C is unaware of the RTS transmission and performs directional transmission to the direction of A. Node A, which is unable to sense the transmission from C, will proceed on CTS transmission and collision will occur since A and C are DD-neighbors that point to its other. In a separate case, node D, that is also unaware of the RTS transmission, performs directional transmission to the direction of A. Node A will sense the transmission from D and abort its CTS reply. However, this is a wrong decision since the transmission from D will be suppressed from A's side lobes and will not affect the nodes A and B communication. If directional channel sensing takes place at node A, these two events will be avoided. However, node A is unable to sense the transmission from node F that will create interference to node B. This could have been avoided by omni-directional carrier sensing.

*CTS Transmission*: By sending a directional CTS packet the potential receiver will inform the nodes towards its main-beam direction for the forthcoming data trans-

mission. However, this does not solve the future hidden terminal problem from the directions of its side lobes. These terminals will be unable to hear the CTS transmission and the will assume that the channel is idle. For example, in Fig. 11.4, node A sends a directional RTS, and node B replies with directional CTS. In this way, node D that is willing to transmit toward the direction of B (and therefore performs directional sensing) will be able to sense the CTS. Note that node D has not sensed the directional RTS from A. Node C, on the other hand, will sense the RTS from A but not the CTS from B and it may assume that the RTS was unsuccessful and the channel is free. Omni-directional CTS transmission does not solve the problem either since the adjacent nodes sense omni-directionally the channel and therefore only the OO-neighbor may sense the CTS transmission (OD-neighbors may be able to sense the CTS transmission only if they point towards the direction of the CTS transmitter). For example, node D (in Fig. 11.4) will not be able to sense an omni-directional CTS transmission from B.

*Data and Acknowledgment Transmission*: Data transmission and positive or negative acknowledgment from the receiver have to take place in a directional way.

*Different Antenna Beamforming Patterns*: In a practical mesh network, it is possible that not all of the wireless transceiver will be equipped with multiple antennas. Therefore, some of the nodes may be able to perform only omni-directional transmission and reception. Furthermore, the radiation patterns (beam-width and directivity) may be different in each transceiver depending on the available number of antenna elements. All these issues have to be also taken into account in the design of realistic medium access control protocols.

As a conclusion, hidden terminal problem is a highly important and still open research issue for nodes communication in a wireless mesh network that is complicated by the introduction of directional transmissions. Traditional RTS/CTS techniques are proven to be inadequate for directional transmissions. Medium access control schemes may have to allocate considerable amount of their wireless resources to mitigate this problem and exploit the benefits of beamforming. Therefore, very careful consideration must be taken into account in the design of higher layer protocols for wireless mesh networks.

### 11.4.3 Congestion Control: Hidden Terminal vs. Deafness Fairness

From the above discussion it is clear that packet collision and deafness are two different issues that cause a node to abort its transmission and increase its back-off period. However, these two events have to be handled in a different way. For instance, the fact that a receiver is deaf towards a specific direction does not imply that the network is highly congested; therefore, an unsuccessful transmitter should not increase its back-off period because this will highly decrease its probability to win the channel contention when the channel becomes finally available. On the other hand, a series of unsuccessful transmissions due to collision implies heavy traffic condition that has to be handled by appropriate congestion control.

Therefore, a novel mechanism has to be defined that differentiate these two events and take appropriate action whenever each of them takes place. Moreover, most of

the existing back-off algorithms have been designed and optimized for an omni-directional network model. Its performance may have to be reconsidered for a directional antenna system.

### 11.4.4 Directional NAV

Another issue that has to be taken into account when directional transmission is employed in a mesh network is the modification of the *Network Allocation Vector (NAV)*. Sensing a transmission from a specific direction does not necessarily imply that a node has to defer its own transmission and modify its back-off timer (as it happens in the omni-directional case). On the contrary, if a node is able to resolve the angle of arrival and deduce that its transmission does not interfere with the ongoing communication, it should presume that the channel is idle. Therefore, a node needs a table to keep track of the directions and the corresponding durations towards which a node must not initiate a transmission, i.e., *Directional NAV (DNAV)*. The continuous update of this table with the right information in order to keep neighbors silenced towards the right direction during a transmission is important both for dealing with the hidden terminal problem as well as for the spatial reuse.

### 11.4.5 MIMO Related Issues

Multiple-input multiple-output (MIMO) links have been seen in the previous section to provide high spectral efficiency in rich multi-path environments through multiple spatial channels without additional bandwidth requirements. This enormous spectral efficiency is obtained for a single link with no external interference. In a wireless mesh environment though, there will be channel reuse and therefore co-channel interference from other APs transmissions. Recent research results have shown that co-channel interference can seriously degrade the overall capacity when MIMO channels are used in a cellular system [29].

Moreover, it has been proven that for flat Rayleigh fading channels, with independent fading coefficients for each path, it is possible to achieve higher capacity by reducing the number of MIMO streams. More specifically, for a system with $n$ receive antennas, $m$ transmit streams and $k$ interfering streams, all the $m$ transmit streams can be isolated and decoded successfully as long as $m + k \leq n$. A group of $m$ antenna elements will be used for data reception while the remaining $n - m$ elements are used to null out the interfering streams. The best performance is achieved when all the degrees of freedom of the MIMO channel are used, i.e., $m = n - k$. On the other hand, if the incoming streams are more than the receiver antenna elements (i.e., $m + k \geq n$), it may not be possible for the receiver to decode any of the desired signal streams if the excess streams degrade the overall SINR below a threshold. It must be noted here that if the interfering ($k$) streams are far weaker than the desired ($m$) streams, it may be possible to decode the desired streams (even if $m + k \geq n$) given that the SINR is above the required threshold.

These observations can be directly applied to the design requirements of multiple access schemes for wireless mesh networks with MIMO channels. Moreover, it has

been shown [20] that for multi-user communication, the channel capacity can be achieved by letting all the users simultaneously transmit and jointly decoded by the receiver rather than organize orthogonal channel access. However, the total number of possible simultaneous transmissions that can be decoded is limited by the number of the antenna elements at the receiver end. Given $n$ receiver antennas, a rule of thumb is to have groups of $n$ users transmit simultaneously and schedule different groups in an orthogonal way (e.g., TDMA).

In a wireless mesh network where multiple antennas are deployed at each AP a transmission can occupy multiple streams depending on the number of transmit antenna elements. In this case, it is more appropriate if a transmitter chooses only a subset of the strongest streams and distributes its power (e.g. using different water-filling techniques) over these streams instead of using the maximum number of streams for transmission [9]. The gain (known as stream control gain) is twofold: only the best channel modes (streams) are used for transmission while on the other hand multiple transmissions take place simultaneously which leads closer to capacity achievement as discussed above. The optimal number of simultaneous transmissions depends on the number of antenna elements used in each transceiver subset, the number of antennas at the receiver end and the number of possible interfering streams.

However, this requires perfect channel state estimation at the transmitter end in order to choose the strongest channel modes. At the receiver side, MMSE is the optimal compromise between maximizing the signal strength from the user of interest and suppressing the interference from the other users. Even better performance can be achieved from an MMSE with successive interference cancelation (MMSE-CSI) receiver.

All these make clear the paramount importance for novel medium access control and routing schemes that exploit these new communication opportunities provided by MIMO channels. For instance, the MAC scheduler should be able to allocate appropriate number of streams per transmitter-receiver pair in a way that a receiver is not overwhelmed by extended number of transmissions. This gives a rise to new kind of hidden and exposed terminal issue. Optimal resource allocation between data and feedback channel must be also performed (for instance, this information can be included in RTS and CTS packets [9]) so that always the best channel models are used. Here, a fairness model (e.g. proportional fairness) should be used in conjunction with the medium access control scheme so that weaker links in the mesh network are not starving. Moreover, appropriate power control schemes are needed so that interference suppression can be performed in the receiver side in conjunction with the stream control. The multiplexing vs diversity trade off must be also taken into account since channel reliability and delay limitations are as much important as high throughput for QoS constrained applications. Finally, QoS routing schemes should include these MIMO parameters (degrees of freedom, stream quality, and multi-stream interference) in their utility functions during route discovery and maintenance.

### 11.4.6 FCC Regulations

In this point we would like to briefly discuss another emerging issue regarding the maximum allowed transmitting power for directional transmissions if the wireless mesh network operates in unlicensed frequency bands. For example, U.S. Federal Communications Commission (FCC) regulations extended the total effective radiated power in unlicensed radio bands from 30dBm for single antenna systems, to 36dBm for beamforming systems. Most of the academic works on MAC and routing schemes for mesh networks do not take into account this parameter and they assume much higher directive gain (compared to the omni-directional transmission) in order, for example, to decrease the number of hops in multi-hop transmissions or increase the network connectivity. However this is a realistic and highly important issue that must be considered in the design of practical wireless mesh networks.

To summarize this section, it is of paramount importance the introduction of novel medium access control and routing schemes that exploit the benefits of multiple antenna deployment on the receiver and/or the receiver side of a wireless mesh network. Without careful consideration in the design of higher layers the usage of smart antenna techniques can have negative impact on the overall mesh network performance. This is why recently the design and performance analysis of medium access control and routing schemes with multiple antennas have attracted high research interest and some of them will be presented in the next section. A cross-layer approach must be taken since most of the issues cannot be handled by individual layers.

## 11.5 Smart Antenna for Scheduling

In the following we describe a number of proposed medium access control schemes with multiple antenna arrays for wireless mesh networks. These protocols are exploiting the benefits of multiple antenna systems while at the same time are trying to overcome the aforementioned problems and challenges that multiple antenna techniques will impose to higher layers. We initially present some CSMA based MAC extensions of the popular IEEE 802.11 protocol that take into account the directional transmission property of smart antennas. We describe a TDMA based scheme with directional antennas and we conclude this section with the description of a couple of MAC protocols that exploit the diversity and multiplexing gain of MIMO systems.

### 11.5.1 Directional-MAC (DMAC)

Directional-MAC (DMAC) [30] is a scheme similar to IEEE 802.11 adapted to the use of directional antennas. The popular 4-way handshake CSMA/CA is used for channel reservation, data transmission and acknowledgment (for more information on the 4-way handshake CSMA/CA and IEEE 802.11 protocol see [27]). The RTS and CTS packets are transmitted directionally while the idle transceivers listen to the channel omni-directionally. In this way a potential receiver is able to estimate the

DoA of the RTS packet and set the direction of the CTS accordingly. More specifically, the DMAC scheme operates as follows:

*RTS Transmission*: Under the assumption that a transmitter knows the location of the receiver, a source performs directional physical carrier sensing towards the direction of the receiver. If the channel is sensed idle the source checks the back-off counter in the Directional Network Allocation Vector (DNAV) where the virtual carrier sensing status for each DoA is maintained. If the back-off counter counts down to zero, the RTS packet is directionally transmitted.

*RTS Reception and CTS Reply*: An idle transceiver is sensing the channel omni-directionally. In this way, the receiver of an RTS packet should be able to determine the DoA of the incoming signal (different DoA estimation techniques such as ESPRIT [13] and MUSIC [14] can be used). If transmission is permitted (both directional physical and virtual carrier sensing) towards the direction of source, the receiver beamforms towards the direction of the source and continues to sense the channel for SIFS time slots. If the channel remains free for the duration of SIFS, the receiver replies with a CTS packet, otherwise, if the carrier is sensed busy, the CTS transmission is cancelled (similar to 802.11 [27]). All the other nodes (except the potential receiver) that are receiving the RTS packet update their respective DNAV tables with the transfer duration specified in the RTS packet. This prevents them from transmission in a certain range towards the reverse direction of the DoA of the receiving RTS.

*CTS Reception and DATA/ACK Reply*: The source node continues to beamform to the direction of the receiver, waiting for the CTS packet reply. If the CTS packet is successfully received within the CTS-timeout duration the DATA transmission starts. Both transmitter and receiver have their beam shifted towards the direction of each other. Upon the successful completion of DATA transfer the receiver replies with an ACK. If the CTS packet is not received within the CTS-timeout duration the transmission is cancelled and the source reschedules the transmission of RTS according to the updated back-off counter. All other nodes that overhear the CTS, DATA and ACK packets update their respective DNAV tables with the directions specified in these packets.

DMAC is a simplified approach that manages to exploit (up to certain point) the beamforming capabilities of multiple antenna transmitters in a wireless mesh network. Nevertheless, DMAC has failed to address the deafness problem while hidden terminals still exit as a result of the omni-directional channel sensing. Moreover, DMAC is unable to establish direct communication between DD-neighbors. See [24] for a more thorough discussion on the problems with DMAC.

### 11.5.2 Multi-hop RTS MAC Protocol (MMAC)

Multi-hop RTS MAC (MMAC) protocol [23] is an extension of the basic DMAC protocol described before. MMAC attempts to exploit the extended transmission range property of directional antennas. If both transmitter and receiver are pointing their beams to each other the communication range can be significantly extended.

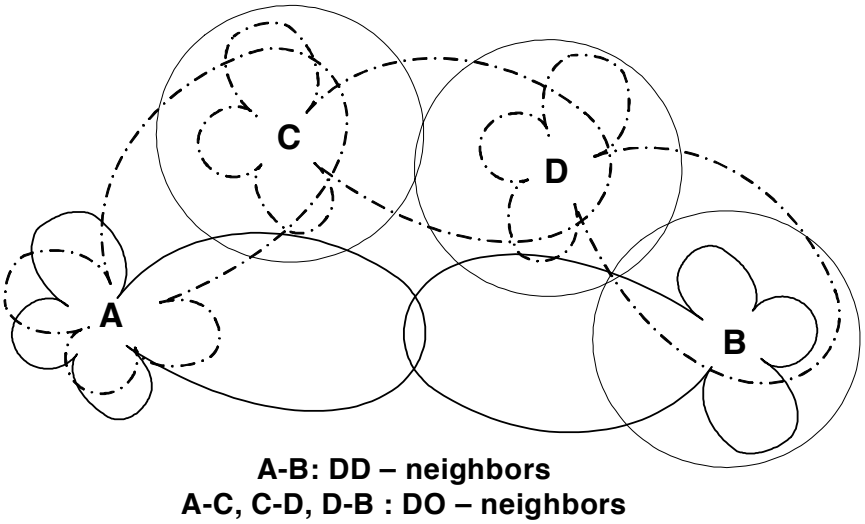



**Fig. 11.5.** DD-link activation in MMAC protocol.

DMAC has failed to address this property since the idle transceivers listens omni-directionally for new transmissions. This is highly important since it can reduce the number of hops between source and destination and reduce the end-to-end delay in multi-hop communications and increase the spatial reuse factor.

Although two DD-neighbors can communicate with each other directly, they need somehow to coordinate their beams to point to each other's direction. This is the main motivation of MMAC protocol that attempts to find an alternative DO-neighbor route for signaling exchange between the two DD-neighbors to coordinate their directions of transmission. To achieve this, an RTS packet has to be forwarded from the DD-source through multiple DO-neighbors to reach the DD-destination while the DD-source is inactively waiting for CTS reply. For instance, in Fig. 11.5, nodes A and B are DD-neighbors that cannot communicate unless they point each other. An RTS packet is sent through the A-C-D-B route to inform B for A's intention. The multi-hop RTS transmission and CTS, DATA and ACK packet exchange mechanism is described in the following:

*RTS Transmission:* MMAC protocol does not provide any neighbor discovery phase. Therefore, it is anticipated that each node have knowledge of the position of all its DD and DO-neighbors. Moreover, it is assumed that a module running above the MAC layer is capable of deciding the appropriate communication scheme (e.g. sometimes it may be more appropriate to use multi-hop DO-route rather than direct DD-transmission) and finding the optimal DO-route to a DD-neighbor.

The source performs directional physical carrier sensing towards the direction of the potential DD-neighbor receiver. If the channel is sensed idle and also the DNAV allows a transmission to this direction, the source is sending an RTS packet towards

the DD-receiver. There is a high probability that this packet may not reach the destination DD-neighbor since this node may be in omni-mode or has shifted its beam to different direction. The aim of such a transmission is to reserve the channel in the region between the DD-neighbors rather than to successfully deliver the RTS packet. Nodes in this region that overhear the RTS transmission will update their DNAV table to defer any transmission towards the directions of both the DD-neighbor nodes for a certain time. This time duration is specified in the RTS packet and is equal to the time required for the RTS packet to reach the DD-receiver plus the CTS packet transmission time (these time durations are described in the following).

If the DD-receiver happens to be beamformed to the direction of RTS, it may directly reply with a CTS packet and the DD-neighbors can proceed on the DATA/ACK phase. Otherwise, the DD-transmitter constructs a special type of RTS packet (called forwarding RTS) that is delivered to destination over multiple DO-hops. This packet contains the information of the DO-neighbors in the route from source to destination. None of the nodes modify their DNAV tables on receiving or overhearing the forwarding-RTS packet. The forwarding-RTS packet gets highest priority for transmission (it does not involve any back-off) while the nodes in the forwarding route simply drop the RTS if they are busy. This implies that the time required for a successful RTS transmission is a constant, known *a priori*.

*RTS Reception and CTS Reply:* In the DD-receiver side two events may happen. If this receiver happens to be beamformed to the direction of the DD-transmitter it can receiver the direct RTS packet. In this case it can directly reply with a directional CTS packet such that the multiple RTS forwarding procedure will be avoided. (Note that, in [23] it is not clear when such an event may happen, since whenever a node is idle and able to receive it will be in omni-mode by default. Nevertheless, this can happen in case the DD-neighbor pair wants to extend their ongoing communication). In the second event, on receiving the forwarding-RTS packet, the DD-receiver proceeds on virtual and physical carrier sensing toward the direction of the DD-transmitter for SIFS time slots (similar to DMAC) and continues with CTS directional transmission.

*CTS Reception and DATA/ACK Reply:* Upon the reception of the CTS packet, the DD-link has been established and the DATA transmission starts. If the DATA transmission finishes successfully, the DD-receiver acknowledges the DD-transmitter with an ACK packet. If no CTS packet received during the CTS-timeout duration the transmission is aborted and the DD-transmitter reschedules the RTS transmission according to its directional back-off timer. Nodes that overhear the CTS and DATA packets update their DNAV with the duration specified in the packets.

MMAC can significantly increase the spatial reuse in a wireless mesh network. Simulation results for random topologies [24] showed that MMAC outperforms DMAC and IEEE 802.11 with omni-directional transmissions. The performance of MMAC is expected to be even better in the case of wireless mesh networks where the topology is more structured and static. Nevertheless, the multi-hop forwarding of RTS packet highly increase the probability of packet collision or drop (since back-off is not used) especially when the traffic demand in the network increases. This can offset the advantages of utilizing DD-links when using MMAC as it increases the average end-to-end delay when compared to DMAC.

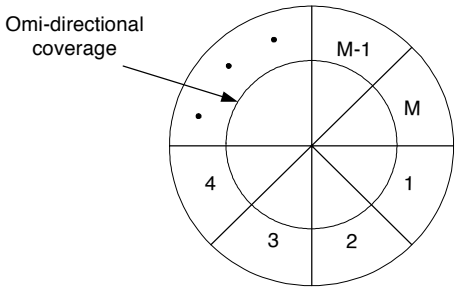



**Fig. 11.6.** Circular directional RTS transmission.

### 11.5.3 CRTS and CRCM

Circular RTS (CRTS) [31] is a simple implementation protocol based on the concept of the IEEE 802.11 but it uses only directional transmissions. It assumes antenna with predefined number if beams i.e., switched beam antenna, that cover all the area around the transmitter. In this scheme the RTS is transmitted in a circular way so that it covers the entire azimuthal-angle domain. For instance, in Fig. 11.6, a nodes with M predefined number of beams starts sending its RTS with beam 1. Shortly afterwards it turns its transmission to beam 2 and transmits the same RTS packet and so on until it covers the whole are by transmitting with beam M. The RTS contains the direction and duration of the intended four way handshake (similar to 802.11).

At the end of the RTS circulation, all the neighbors of the transmitter are informed about the intended transmission and after executing a simple algorithm [31] decide if they will defer their transmission in the direction of transmitter or receiver. On the other hand, the transmitter hears the channel omni-directionally to receive the CTS reply within a predefined period. Note that the receiver has to wait an appropriate amount of time until the transmitter finishes the RTS circulation and switches to receiving mode before it proceeds to its CTS reply.

Circular RTS and CTS MAC (CRCM) [32] is an extension of CRTS that further improves the robustness of medium access control by introducing a combination of circular transmission of RTS and CTS messages. In CRTS, not all of the receiver's neighbors are made aware of the pending DATA and ACK transmissions and therefore it does not solve the hidden terminal problem in receiver's neighborhood. In particular, these nodes can initiate transmissions that can cause a collision during the ACK reception at the transmitter.

In order to tackle this problem the CRCM algorithm introduces an efficient mechanism for the directional transmission of CTS. The circular RTS transmission is the same as CRTS, however, the receiver after replying with a directional CTS to the transmitter will further transmit directional CTS messages towards the unaware neighboring nodes (i.e., those nodes that are in the coverage range of the receiver but not in that of the transmitter). For example, in Fig. 11.7, node B avoids the trans-

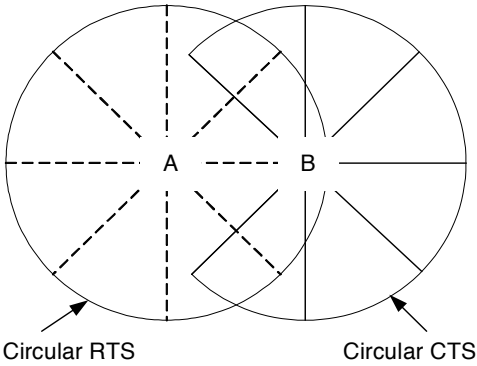



**Fig. 11.7.** Circular directional RTS and CTS transmission.

mission of CTS to the directions that have been already covered by the circular RTS message. The receiver's neighbors execute the same algorithm [32] as did the transmitter's neighbors in order to decide on whether or not to postpone their transmission towards the sender-receiver pair.

CRCM algorithm gives a solution to the hidden terminal problem due to asymmetry in gain, arising due to the deployment of directional antennas in wireless ad hoc or mesh networks. While power consumption because of the extensive circular packet transmission is not an important issue in mesh networks, the time spent for these transmissions can be proven crucial for several delay-sensitive applications.

### 11.5.4 ToneDMCA

ToneDMAC [24] is a medium access control scheme that tries to alleviate the impact of deafness problem while retaining the benefits of directional beamforming. Similar to previous directional MAC protocols, ToneMAC adopts the CSMA/CA principles of IEEE 802.11 and combines it with switched beam antenna transceiver and a tone-based mechanism. The main idea is that both transmitter and receiver, after the completion of their communication, transmit omni-directionally out-of-band tones to differentiate deafness from congestion. In this way, the backlogged nodes can deduce deadness as the cause of their previous failure (and not congestion) and adjust their back-off interval accordingly.

At the beginning, the transmitter and receiver exchange directional RTS and CTS packets without trying to inform their "omni-directional" neighbors of their intended communication. If the initial handshake is successful, data and ACK packets transmission follow in a similar directional way. After the completion of their dialogue, both nodes omni-directionally transmit out-of-band tones to notify their neighbors that it was them that have been engaged in communication over the recent past.

The intended transmitter beamforms in the direction of its intended receiver and performs physical carrier sensing towards that direction. If the channel is sensed idle, the transmitter randomly chooses the time it will transmit within the back-off window interval $[0, CW_{min}]$. Then it switches to omni-directional mode and senses the channel while performing the back-off counting. If a signal is sensed, it performs azimuthal beam scan to determine its DoA. If the signal arrives from a direction different than the direction of its potential receiver, it continues with the countdown. Otherwise it stops the counter until the channel is sensed idle again. Note here that since this node is in sensing mode, it may happen to receive a RTS packet that is intended for itself. In this case, it may choose to abort its own transmission and reply with a CTS packet, alleviating the possibility of deafness and deadlock.

In ToneMAC, the channel is divided into two sub-channels: a data channel where RTS, CTS, data and ACK packets are transmitted, and a narrow control channel where the busy tones are sent. The busy tones do not contain any information (e.g. sinusoids with sufficient spectral separation). These tones can only be detected (through energy estimation) but the receivers will not be able to determine the sender of the tones. To solve this issue, ToneMAC proposes a simple hash function that allocates tones of different frequencies and time durations to different nodes according to node's identifier. If a node $i$ can choose a tone of frequency $f_i$ from a set of $K$ frequencies, and an integer time duration $t_i$ from an interval $[t_{min}, T]$, a simple hash function is used to assign a tuple $(f_i, t_i)$ to node $i$ such that

$$f_i = (i \mod K) + 1 \tag{11.1}$$

$$t_i = (i \mod (T - 1)) + 2. \tag{11.2}$$

The transmit power of the tones is increased such that the omni-directed transmission range will be equal to the range of directional transmission. (Please note that even this is what is mentioned in [24], smaller transmission power will be appropriate since the busy-tone has only to be sensed but not decoded [12]). Of course there is a small probability that two or more nodes have the same $(f, t)$ signature or they may have the same frequency and different tone duration but their tones overlap in time. This probability is quite small since these nodes may be located far away from each other or they point in different directions; also their randomized back-off counters further reduce the probability of simultaneous tone transmission (for more information on this issue see [24]).

ToneDMAC comprises a practical solution for CSMA based mesh networks with directional antennas due its ability to reduce channel-idle time and resolve deafness deadlocks. Nevertheless, since tones are transmitted on a narrow bandwidth channel, and the duration of transmission is short, tones may be detected partially, or may not be detected at all. Multi-path effects can also cause a tone to arrive from a different direction than the known direction of a neighbor. Clearly, both these effects can cause a node to misclassify the cause of transmission failure.

### 11.5.5 Directional Transmission and Reception Algorithm (DTRA)

DTRA [33] is a TDMA-based MAC algorithm for load-dependent slot reservation in a wireless ad hoc network with directional antennas. The main characteristics of DTRA are that (1) it provides fully directional transmission/reception, (2) it is distributed, (3) it provides dynamic on-demand slot allocation and reservation to different links, and (4) it includes power control. Its TDMA-based reservation scheme makes it appropriate for quality-of-service (QoS) support in wireless ad hoc and mesh networks. DTRA requires that each node is equipped with one transceiver with a steerable antenna and all nodes are synchronized. Each node can rapidly switch between transmitting and receiving mode (half duplex). In order to communicate, two nodes must point their beams at each other at the same time and be in a complementary transmit-receive mode. The timeline is divided into frames that each of them comprises three sections dedicated to neighbor discovery, data reservation and data transmission.

*Neighbor Discovery Phase:* During neighbor discovery phase a node performs a scan by transmitting a sequence of advertisements in each possible direction. A three-way handshake is used, where the receiver node of an advertisement replies with its own advertisement and expects to receive an acknowledgment in return within a short time interval. During the three-way handshake, the nodes can exchange transmit power level information and also redefine their directional information. Moreover, they make an agreement on the future mini-slots that they will listen to each other in the reservation phase (a detailed description of the handshake algorithm and the information exchanged is given in [33]). These agreements are valid until the time the two nodes detect each other again. The actual number of scans needed by a node to discover all of its potential neighbors depends on the characteristics of the achievable network graph, its beam-width, and the algorithm used in each node for their mode (scanning or listening) decision. A deterministic mode selection algorithm was proposed so that two neighbors can be detected by each other in at most $log_2 N$ scans, where $N$ is the maximum number of nodes in the network. A random mode selection algorithm was also proposed in [34] where a node decides whether to be in scan or listen mode with equal probability independent of the decisions made in previous slots.

*Slot Contention and Reservation Phase:* In the reservation phase, reassurance of two nodes connection and reservation of data slots will take place. Two nodes will re-detect each other at the predefined mini-slots by pointing in the direction agreed on during neighbor discovery. A three-way handshake will take place similar to the discovery phase where the two nodes negotiate who is going to transmit/receive and agree on the mutual available slots to be reserved for data transmission in the next phase. It is possible that none of the nodes desire to make a reservation or the mutual available slots are not sufficient for their communication (e.g. QoS requirement are not satisfied).

*Data Transmission and QoS:* In order to provide QoS support, three priority queues for each neighbor are defined and the available slots in the data transmission phase are ranked with a priority metric. When making a reservation, if there

are not sufficient slot for high-priority traffic, slots allocated to lower priority can be asked to be reallocated. To ensure fairness, a threshold on the maximum number of slots for each priority class should be set.

One of the main advantages of DTRA protocol is that it takes advantage of the fully directional communication (both directional transmission and reception all the time) capability of a multiple antenna system. Moreover, its slot reservation scheme makes it appropriate to a mesh network with various QoS constrains and demands. One of the main disadvantages is its delay performance for low traffic (e.g., compared to omni-directional IEEE 802.11) because of the fixed resource allocation phase. Such delays will be accumulated in each hop in a multi-hop scenario. Since it is based on directional transmission, the performance of DTRA will highly degrade in multi-path communication environment without a dominant LOS component.

### 11.5.6 SD-MAC

SD-MAC [35] is an IEEE 802.11 based medium access control scheme that exploits the spatial diversity gain of MIMO systems, given that the wireless transceiver are employed with multiple antennas. In order to achieve full-order spatial diversity, space time coding is used on the transmitter end. CSI knowledge is required at the transmitter while channel gains at the receiver are obtained by using preamble symbols. SD-MAC is based on the RTS/CTS mechanism of the IEEE 802.11 distributed coordination function (DCF). More specifically, a source node performs channel virtual carrier sensing similar to 802.11 and physical carrier sensing by using all its antenna elements. If NAV is empty and channel is sensed free, the RTS packet is transmitted in a default rate using space-time coding in order to exploit the transmission diversity. The destination node receives the RTS and uses the preamble symbols to perform channel estimation before it decodes the space-time encoded packet. Other nodes that overhear the RTS packet decode the transmission duration and update their NAV tables. The destination node replies with a space-time encoded CTS packet that contains the rate control information for the following DATA packet based on the channel estimation. The source node receives the CTS packet, adapts its transmission data rate according to the information passed by the CTS and transmits the multi-rate DATA packet. The destination replies with a default-rate ACK packet to confirm the data reception.

While SD-MAC scheme exploits the spatial diversity of the MIMO system, it fails to exploit multiplexing gain. This could be achievable since the RTS/CTS packets can be used as a feedback channel for CSI information at the transmitter. The diversity can improve significantly the channel reliability and increase the transmission rate indirectly. On the other hand, diversity can also increase the transmission range for the same data rate. This can have a great impact on the routing scheme since it can reduce the number of hops and decrease the corresponding end-to-end delay. However, a higher transmission range will increase the area of coverage which can lead to delays due to contention and packet collisions. The impact of SD-MAC on the routing in terms of the optimal hop distance was analyzed in [35].

### 11.5.7 Stream Controlled Medium Access (SCMA)

Both a centralized and distributed versions of Stream Controlled Medium Access were proposed in [9]. Since the main focus of this chapter is on distributed algorithms for mesh networks, only the latter version will be presented here. The main objective of SCMA is to maximize the network utilization subject to a given fairness model. More specifically, SCMA tries to leverage the benefits of stream control (transmission on a subset of the strongest streams) and partial interference suppression (use the remaining streams for interference suppression) in a distributed way where all the available degrees of freedom are shared between the mesh network transceivers. The main components of the SCMA algorithm are presented in the following:

*Node Coloring:* Initially the distributed SCMA protocol performs identification of bottleneck links, i.e., the links that belong to multiple contention regions in the network. These bottleneck links are colored "red" while the remaining nodes of each contention region are colored "white". The red links are scheduled in a non-stream controlled manner (operate on all available steams) while the white links are based on pure stream control. This is due to the fact that the red links are consuming resources from multiple contention regions that otherwise can operate simultaneously on all their available resources. Therefore it is preferable that the red links are scheduled independently (for some examples on how this differentiation will improve the overall resource allocation together with the complete coloring algorithm description, see [9]).

*Contention and Channel Access:* After the nodes have chosen their color, they move to the channel contention phase that depends on their color. This includes four modes: (1) $No\_Contend$,(2) $Contend$,(3) $Acquire$ and (4) $Sched\_White\_Links$. Every node is initially in the $No\_Contend$ mode. Whenever a node has a packet to transmit, it moves to $Contend$ mode with probability $P_{new} = P_{old}$ ($P_{old}$ is a network parameter that represents the persistence value; the entire network adaptation progress is based on this value). In *Contend* mode, a node chooses a waiting time (in number of slots) uniformly distributed from the interval $(0, B)$ after which it senses if the channel is busy. The busy state of the channel here corresponds to a lack of available degrees of freedom (streams) in the channel around the transmitter and the receiver. If the channel is sensed busy, the node aborts its transmission and readjusts its persistence to $P_{old} = (1-\beta)P_{old}$. Moreover, if the it is a white node, it further updates its $P_{new}$ value to $P_{new} = (P_{old}K_{old})/K_{new}$. Same readjustment is performed if the node faces or detects a collision.

If the channel is sensed idle, the node proceeds to $Acquire$ mode where the actual data transmission takes place. Every node in the two-hop range away from the transmitter automatically expends the appropriate number of resources (antenna elements) to suppress the interference from this transmission. At the end of a transmission, all the neighbour nodes having a packet to transmit increase their persistence $P_{old} = P_{old} + \alpha$, while the white nodes further update their $P_{new} = (P_{old}K_{old})/K_{new}$. Typical values for $\alpha$ and $\beta$ are 0.1 and 0.5, respectively [9]. In order to determine the resource availability a node has to estimate each time the amount of remaining streams of every node in its two hop neighbourhood by listening to the con-

trol (RTS/CTS) packets transmissions. Therefore the reception range of the control packets has to be extended by a factor of two. This can be achieved by transmitting multiple copies of the control packets on at least four streams (double range can be achieved with 4 antenna elements and a path-loss exponent of 4). Since CSI is not available at the receiver in this stage, space-time block codes can be used to exploit the transmit diversity gain of MIMO for range extension.

*Coordinated Scheduling:* The first white node in a clique, which gains access to the channel, coordinates the other white nodes to transmit in the same slot using their own estimated fair share. This is the $Sched\_White\_Links$ mode. This is achieved by the introduction of a flag in the RTS/CTS packets. All the white links that have a packet to transmit schedule themselves in the same slot, irrespective of whether they contend for channel access in that slot or not. However, all the transmitting white links (except the initiator of the coordinating scheduling) will still have to update their persistence to $P_{old} = (1 - \beta)P_{old}$.

SCMA algorithm comprises a novel technique to exploit the propitious characteristics of MIMO links in wireless mesh networks. SCMA can improve the aggregate network throughput and improve the fairness [9] compared to CSMA/CA($k$), i.e. conventional CSMA with transmission over all the $k$-streams. However, as the number the node density increases, more independent contention regions will overlap, and as a result, more nodes will be colored red and the SCMA network performance will converge to CSMA/CA($k$). Moreover, further investigations are needed on how the power of adjusted interfering streams does affect the ongoing data stream transmissions.

### 11.5.8 Conclusions on Scheduling with Multiple Antennas

In this section several distributed medium access control protocols with multiple antennas deployment have been presented and their advantages and drawbacks have been discussed. The majority of these schemes ( [30] [23] [31] [32] [24]) are extensions of the popular IEEE 802.11 protocol, therefore, are based on random channel access that makes them inappropriate for strict QoS constraints. DTRA [33] protocol provides slot reservation that can promise QoS at the price of relatively high delays for low traffic networks as discussed before. All these protocols are based on beamforming techniques. SD-MAC [35] and SCMA [9] on the other hand, are exploiting the diversity and multiplexing gain respectively of the MIMO channel in order to increase channel reliability and data throughput.

The overall mesh network system performance can be further improved if opportunistic transmissions are considered. Recent work on opportunistic scheduling [36] with omnidirectional transmissions has shown that by using appropriate utility functions considerable opportunistic gain can be achieved, while at the same time the generated interference is reasonably temporal-correlated. This is an important property that ensures satisfactory channel prediction for better distributed power control and scheduling performance. Opportunistic scheduling schemes combined with the aforementioned multiple antenna techniques is an unexplored and promising area of

high research interest and potentials (e.g., it is part of the research agenda for the MEMBRANE project [3]).

## 11.6 Smart Antennas for Routing

While multiple antenna techniques have been widely analyzed from a MAC perspective, their usage and impact on network layer and more specifically their interaction with routing has not received much research attention. Moreover, the research community interest over the last decades regarding wireless routing has been only concentrated on omnidirectional transmissions. In the following we briefly demonstrate and discuss a number of proposed routing schemes that take into account multiple antenna techniques (mainly for directional transmission).

The impact of smart antennas on QoS routing for multi-hop wireless networks was evaluated by simulations in [37] as an extension of the Wireless Fixed Relay routing (WiFR) [38] protocol. However, the analysis is based on a mathematical programming model, and no routing algorithm was defined in terms of signaling that has to be exchanged through the route and the required cross layer interaction in order to solve the aforementioned problems related with directional transmissions.

The routing improvement using directional antennas in adhoc networks was demonstrated in [39] where two techniques are proposed to a) bridge permanent network partitions and b) repair routes in use in case of link breakage by using directional transmissions. The design and evaluation was based on the Dynamic Source Routing protocol [40], an on-demand routing protocol for mobile ad hoc networks, in which, the originator of a packet decides the entire sequence of hops through which the packet is to be forwarded to the final destination. However, this protocol is designed to enhance network connectivity rather than increasing the end-to-end throughput or guarantee quality of service. Therefore, the directionality is used only when two nodes are located far enough for omnidirectional communication (the decision is based on SINR measurements).

Another approach on routing with beamforming was illustrated in [41]. The proposed algorithm is based on the well known Ad-Hoc On-Demand Distance Vector routing protocol (AODV) [42] with directional transmitters and omnidirectional receivers. Nodes are assumed to be equipped with switched beam antennas consisting of $K$ directional non-overlapping beams each of them spanning an angle of $2\pi/K$ radians. For unicast packets, if the destination nodes are located on the direction of the same beam, the transmissions are time multiplexed, while different beams are simultaneously activated if the receivers are located in separate directions. For broadcasting messages, such as *Route Request*, all the beams are activated. However, this work also fails to address the problem of deafness and the beam synchronization for DD-nodes communication.

While the previous work on routing with multiple antennas is mainly focused on using the directionality to increase connectivity, a wireless mesh network must also provide high data throughput and meet the various QoS requirements of the overlying applications. Smart antennas is definitely a key technology that can highly contribute

towards this direction. Therefore, it is of paramount importance to design efficient QoS routing protocols that harmonically coexist with lower layers and exploit the multiple capabilities and benefits that the multiple antenna technologies can provide. For instance, novel routing schemes exploiting the SDMA opportunities of MIMO techniques have to be investigated in conjunction with new utility functions comprising of new interference patterns and increased transmission ranges. Furthermore, algorithms for DD-synchronization and communication through the route have to be designed to reduce end-to-end packet delays.

## Conclusion

The impact of multiple antenna array techniques on medium access control schemes for mesh networks has been analyzed in this chapter. Several multiple antenna architectures and methodologies, such as steered-beam antennas, adaptive antennas, and MIMO coding, have been demonstrated. It has been clear that while these techniques can significantly improve the performance of single-link communications, from a wireless mesh network perspective they can also considerably degrade the overall performance of wireless mesh networks if careful consideration of their interaction with higher layers is not taken into account. Different design challenges, such as deafness, hidden and exposed terminals and MIMO related issues have been discussed. Several proposed MAC protocols for wireless ad hoc and mesh networks with directional antennas and MIMO techniques have been presented and their advantages and weaknesses have been discussed. Finally, we have briefly discussed the lack of efficient routing algorithms for nodes with multiple antennas and the need of routing schemes that exploit the smart antenna's capabilities.

## Acknowledgment

This work was performed for the European 6th Framework MEMBRANE project (IST-4-027310).

## References


1. I. F. Akyldiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, pp. 23-30, vol. 43, no. 9, Sept. 2005.
2. I. F. Akyldiz, X. Wang, and W. Wang, "Wireless mesh networks: A survey," *ELSEVIER Computer Networks*, pp. 445-487, vol. 47, no. 4, March 2005.
3. IST-MEMBRANE: Multi-Element Multihop Backhaul Reconfigurable Antenna Network (www.ist-membrane.org).
4. P. H. Lehne and M. Pettersen, "An overview of smart antenna technology for mobile communication systems," *IEEE Communications Surveys*, pp. 2-13, vol. 2, no. 4, Fourth Quarter 1999.
5. A. Alexiou and M. Haardt, "Smart antenna technologies for future wireless systems: Trends and challenges," *IEEE Communications Magazine*, pp. 90-97, Sept. 2004.
6. M. Webb, M. Beach, and A. Nix, "Capacity limits of MIMO channels with co-channel interference," in *Proc. of VTC 2004-Spring*, 17-19 May 2004.
7. Y. Song and S. D. Blostein, "MIMO channel capacity in co-channel interference," in *Proc. of 21st Biennial Symposium on Communicatios*, pp. 220 - 224, June 2002.
8. S. Bandyopadhyay, S. Roy, and T. Ueda, *Enhancing the Performance of Ad hoc Wireless Networks with Smart Antennas*, Auerbach Publications, 2006.
9. K. Sundaresan, R. Sivakumar, M. A. Ingram, and T.-Y. Chang, "Medium access control in ad hoc networks with MIMO links: Optimization considerations and algorithms," *IEEE Trans. on Mobile Computing*, pp. 350-365, vol. 3, no. 4, Oct.-Dec. 2004.
10. J. H. Winters, "Smart antennas for wireless systems," *IEEE Personal Communications*, pp. 23-27, vol. 5, no. 1, Feb. 1998.
11. J. H. Winters, "Smart antenna techniques and their application to wireless ad hoc networks," *IEEE Wireless Communications*, pp. 77-83, vol. 13, no. 4, Aug. 2006.
12. J. A. Stine, "Exploiting smart antennas in wireless mesh network using contention access," *IEEE Wireless Communications*, pp. 38-49, vol. 13, no. 2, April 2006.
13. R. H. Roy and T. Kailath, "ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, pp. 984-995, vol. 37, no. 7, July 1989.
14. R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas Propagation*, pp. 276-280, vol. 34, no. 3, March 1986.
15. S. Bellofiore, J. Foutz, R. Govindaradjula, I. Bahceci, C. A. Balanis, A.S. Spanias, J.M. Capone, and T. M. Duman, "Smart antenna system analysis, integration and performance for mobile ad hoc networks (MANETs)," *IEEE Trans. on Antennas and Propagation*, pp. 571-581, vol. 50, no. 5, May 2002.
16. S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications (JSAC)*, pp. 1451-1458, vol. 16, no. 8, Oct. 1998.
17. V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. on Information Theory*, pp. 1456-1467, vol. 45, no. 5, July 1999.
18. P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," *in Proc. of URSI International Symposium on Signals, Systems, and Electronics*, Sept.-Oct. 1998.
19. G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, pp. 41-59, vol. 1, no. 2, Fall 1996.

20. D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.
21. L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Information Theory*, pp. 1073-1096, vol. 49, no. 5, May 2003.
22. K. Sundaresan, S. Lakshmanam, and R. Sivakumar, "On the use of smart antennas in multi-hop wireless networks," in *Proc. of BROADNETS*, Oct. 2006.
23. R. R. Choudhury, X. Yang, R. Ramanathan, and N. H. Vaidya, "Using directional antennas for medium access control in ad hoc networks," in *Proc. of ACM MOBICOM'02*, Sept. 2002.
24. R. R. Choudhury and N. H. Vaidya, "Deafness: A MAC problem in ad hoc networks when using directional antennas," in *Proc. of the 12th IEEE International Conference on Network Protocols (ICNP'04)*, Oct. 2004.
25. F. A. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II - the hidden terminal problem in carrier sense multiple-access modes and the busy-tone solution," *IEEE Trans. on Communications*, pp. 1417-1433, vol. COM-23, no. 12, Dec. 1975.
26. P. Karn, "MACA - A new channel access method for packet radio," in *Proc. of the 9th ARRL Computer Networking Conference*, 1990.
27. ANSI/IEEE Std 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layers (PHY) Specifications," 1999.
28. G. Li, L. L. Yang, W. S. Conner, and B. Sadeghi, "Opportunities and challenges for mesh networks using directional antennas," in *Proc. of IEEE WiMesh*, Sept. 2005.
29. M. F. Damirkol and M. A. Ingram, "Stream control in networks with interfering MIMO links," in *Proc. of Wireless Communications and Networking WCNC*, March 2003.
30. Y. B. Ko, V. Shankarkumar, and N. H. Vaidya, "Medium access control protocols using directional antennas in ad hoc networks," in *Proc. of IEEE INFOCOM*, March 2000.
31. T. Korakis, G. Jakllari, and L. Tassiulas, "A MAC Protocol for full exploitation of directional antennas in ad-hoc wireless networks," in *Proc. of ACM MobiHoc*, June 2003.
32. G. Jakllari, I. Broustis, T. Korakis, S. V. Krishnamurthy, and L. Tassiulas, "Handling asymmetry in gain in directional antenna equipped ad hoc networks," in *Proc. of IEEE PIMRC*, Sept. 2005.
33. Z. Zhang, "DTRA: Directional transmission and reception algorithms in WLANs with directional antennas for QoS support," *IEEE Network*, pp. 27-32, vol. 19, no. 3, May-June 2005.
34. M. E. Steenstrup, "Neighbor discovery among mobile nodes equipped with smart antennas," in *Proc. of ADHOC'03*, May 2003.
35. M. Hu and J. Zhang, "MIMO ad hoc networks: Medium access control, saturation throughput and optimal hop distance," *Journal of Communications and Networks*, Special Issue on Mobile Ad Hoc Networks, pp. 317-330, Dec. 2004.
36. Y. Hou and K. K. Leung, "A novel distributed scheduling algorithm for wireless mesh networks," in *Proc. of IEEE Globecom 2007*, Nov. 2007.
37. L. Coletti, D. Ciglioni, A. Capone, and M. Zambardi, "Impact of smart antennas on QoS routing for muli-hop wireless networks," in *Proc. of the 14th IST Mobile and Wireless Communications Summit*, June 2005.
38. A. Capone, L. Coletti, and M. Zambardi, "QoS routing in multihop wireless networks: New model and algorithm," in *Proc. of QoS-IP 2005*, Feb. 2005.
39. A. K. Saha and D. B. Johnson, "Routing improvement using directional antennas in mobile ad hoc netwroks," in *Proc. of IEEE Globecom 2004*, Nov. 2004.
40. D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, ch. 5, pp. 153-181, Kluwer Academic Publishers, 1996.

41. N. Nie and C. Comaniciu, “Energy efficient AODV routing in CDMA ad hoc networks using beamforming,” in *Proc. of IEEE Vehicular Technology Conference (VTC)*, May-June 2005.
42. C. E. Perkins, “Ad hoc On-Demand Distance Vector (AODV) Routing,” RFC 3561, IETF Network Working Group, July 1998.

# 12

# Security Issues in Wireless Mesh Networks

W. Zhang[1], Z. Wang[2], S. K. Das[1], and M. Hassan[2]

[1] The University of Texas at Arlington, USA
{wzhang, das}@cse.uta.edu
[2] University of New South Wales, Australia
{zhewang, mahbub}@cse.unsw.edu.au

## 12.1 Introduction

With recent advances in wireless technologies such as multiple-input multiple-output (MIMO) systems and smart antennas, wireless mesh networks (WMNs) have attracted increasing attention as an alternative for large-scale deployment of metropolitan area wireless networks.

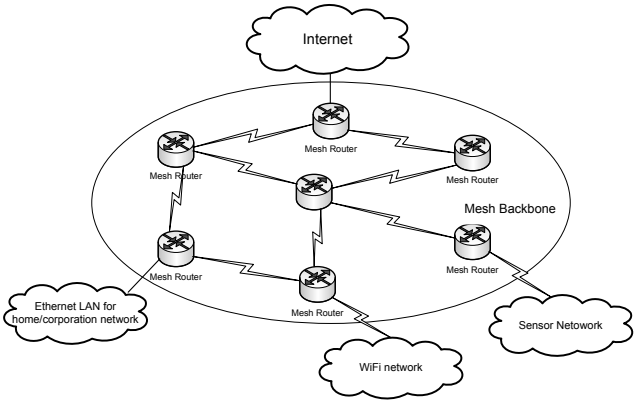



**Fig. 12.1.** Typical infrastructure of WMN.

Fig. 12.1 illustrates a typical architecture of a WMN. The mesh routers constitute a self-configuring, self-healing network backbone. The various types of networks interconnected by the backbone communicate with each other through the

wireless multihop links between the mesh routers. WMNs share some nice features with wireless ad hoc networks, including self-organization and self-configuration. However, since the mesh routers are either static or with minimal mobility, there exists a predictable infrastructure in WMN. Thus, WMNs have the advantage of being extremely easy to deploy and relatively cheap in terms of both infrastructure and maintenance cost.

These desirable features make WMNs an appealing solution for a plethora of applications, such as broadband home networking, community networking, etc. However, there are still several challenges and issues preventing WMNs to be widely deployed in large scales. The first major issue is that the performance (throughput, delay, or packet loss rate) of WMNs drops sharply with increasing number of wireless hops the packets traverse through. The multi-radio, multi-channel technique ( [2], [48]) is being researched to overcome this problem. The second major issue is the lack of an integrated cross-layer solution to provide *security* in WMNs, which has received meager attention in the literature. Clearly, without a well designed security solution, WMNs are vulnerable to various types of internal and external attacks that may cause significant inconvenience to the users and operators.

In this chapter, we will address the security issues in wireless mesh networks. The rest of the chapter is organized as follows. In Section 12.2 we discuss the security goals and challenges posed in WMNs. Section 12.3 surveys and analyzes the applicability of existing security techniques to WMNs. In Section 12.4 we point out the open problems in this area. Finally, we conclude this chapter.

## 12.2 Security Goals and Challenges

For any application (not necessarily on WMNs), the following general goals are desired to ensure security.

*Confidentiality* or *Privacy*: The communication between users must be secured such that the information cannot be disclosed to any eavesdroppers.

*Integrity*: The whole transmission paths must be protected to ensure the messages are not illegally altered or replayed during the transmission.

*Availability*: Applications should provide reliable delivery of messages against denial of service (DoS).

*Authentication*: When a user sends messages, there should be some processes to identify the user to ensure the messages are really sent by the claimed sender rather than fabricated by someone else.

*Authorization*: Before any user performs some tasks, there should be mechanism to ensure the corresponding users have the right to do them.

*Accounting*: When a user is using some services, some process should be able to measure the resources the user consumes for billing information.

Here we assume the existence of upper layer security mechanisms, such as antivirus software and Secure Sockets Layer (SSL) protocol, and focus on additional security challenges posed by the unique features of WMNs.

### 12.2.1 WMN Specific Security Challenges

The shared nature of wireless medium, the absence of globally trusted central controller, and the lack of physical protection of mesh routers pose the main challenges for securing WMNs.

First, like any wireless networks, the shared wireless medium makes it easy for attackers to launch jamming attacks, eavesdrop the communication between the mesh routers and inject malicious information into the shared medium. Given the fact that the correctness of routing messages is fatal to achieve wireless multihop routing in WMNs, the most harmful kind of malicious information is due to the fabricated routing messages.

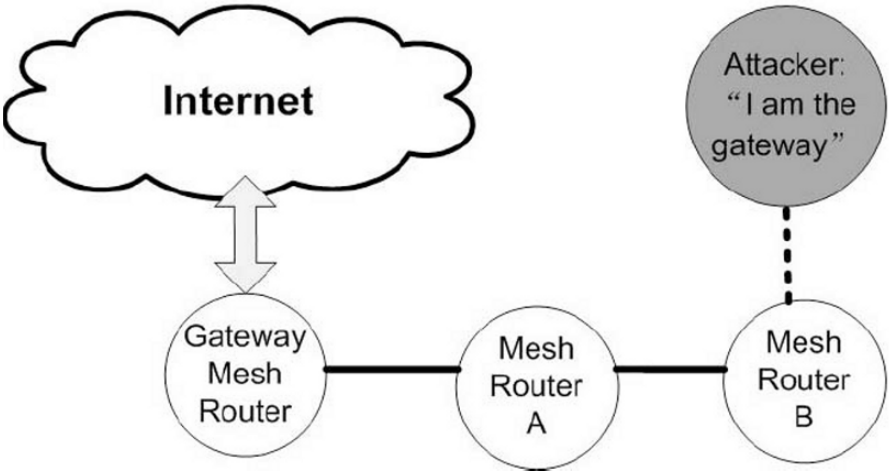



**Fig. 12.2.** The attacker fools mesh router B.

Fig. 12.2 illustrates a simple example of such attacks. The correct route for *Mesh Router B* to access the Internet is via *Mesh Router A* and the *Gateway Mesh Router*, while the attacker fools *Mesh Router B* by broadcasting the message: *"I am the gateway to the Internet."* If *Mesh Router B* could not detect such a message as faulty, it will direct all its Internet traffic to the attacker. Because the wireless medium is open, it is impossible to prevent the mesh routers from receiving such malicious messages. Therefore, an authentication mechanism is essential to distinguish the malicious information from the legitimate information.

Second, an authentication mechanism is usually implemented with the help of *Public Key Infrastructure* (PKI), which requires a globally trusted entity to issue certificates. However, it is impractical to maintain a globally trusted entity in WMNs. The details of authentication challenges are discussed in Section 12.3.1.

Third, the mesh routers are located outdoor, usually on roof tops or traffic light poles. They are not physically protected like the wired routers and wireless LAN access points. Therefore, it is much easier for attackers to capture the mesh routers and

get full control of the device. If a router is fully controlled by attackers, the attacks can be launched from that router and the information sent by the attackers will be regarded as authenticated by other routers. The cryptographic authentication schemes are thus broken and there must be another line of defense behind the authentication protection.

The above major challenges demand a set of cross-layer, self-adapted security mechanisms to protect WMNs.

In the following sections, we will discuss if and how some of the existing security solutions proposed for wireless ad hoc or sensor networks could be employed to protect WMNs by overcoming these challenges.

## 12.3 Security Concerns and Current Countermeasures

While the security of WMNs is a fairly new research topic, there exist several schemes to secure wireless ad hoc networks and wireless sensor networks which share similarities with WMNs to some extent. Let us analyze these solutions and discuss how to utilize them to secure WMNs.

### 12.3.1 Authentication

In wireless networks, authentication is very important because of the shared nature of the wireless medium. Any node, legitimate or malicious, with a suitable hardware device can send data into the network. Verifying that the data received is from a legitimate entity is critical for securing the network. Public key infrastructure (PKI) and certification authority (CA) provide two important mechanisms for authentication.

#### PKI and CA

Authentication is usually realized by implementing PKI based on asymmetric cryptography in which each user has a pair of cryptographic keys: *public key* and *private key*. The public key is widely distributed and known by all the users while the private key is only secretly kept by the user. One property of the pair of keys is that a message encrypted with the public key can only be decrypted with the corresponding private key and vice versa. By exploiting this, authentication can be achieved. For instance, a sender can digitally sign the packets using its own private key before sending them. If the receiver can successfully decrypt the messages with the sender's public key, it is assured that the packets are really sent by the claimed sender rather that someone else.

To check the validity of a digital signature, it is necessary to first verify that the sender's public key does belong to the sender, which requires a Certificate Authority (CA) be involved in the authentication procedure. The CA signs the binding of an entity's identity and its public key with its private key, and issues the signature as the entity's certificate. Any entity can validate the binding of sender's identity and

public key by checking its certificate using CA's public key. A node may update its certificate periodically to reduce the chance of brute-force attack on its private key. So the CA has to stay on-line to reflect the periodically changing certificates. This scheme is based on the following assumptions: (a) the CA's public key is known by every entity in the network, (b) the CA's public key and signed certificates are globally trusted in the network, and (c) the communication channels through which the entities get other's certificate from CA are secure.

However, the absence of pre-established trusted network infrastructure in WMNs obstructs direct application of PKI. This is because it is impractical to deploy a CA that every node can trust and establish a secure communication channel with. A distributed CA scheme is thus required.

**Distributed CA**

An ingenious method is to distribute the functionality of the centralized CA to the whole network by applying *threshold cryptography* [15]. Basically, an $(n, t+1)$-threshold cryptography scheme allows $n$ parties to share the ability to create a digital signature so that $t+1$ parties can jointly generate a valid signature, whereas it is infeasible for at most $t$ parties to do so. The scheme is based on the assumption that the number of compromised parties will never exceed $t$.

In WMNs, if the CA's public key is globally known and its private key is divided into $n$ shares (one share for each node in the network), the threshold signature scheme [62] can be designed so that the certificate of a particular node is signed by combining $t+1$ partial signatures generated by $t+1$ nodes respectively, and the certificate can be verified by the CA's public key which is known by each node in the network.

Fig. 12.3 shows a $(5, 3)$-threshold signature scheme, in which the CA's private key is divided into 5 shares for each node: $s1, s2, ..., s5$. A message $m$ (the identity and public key of a particular node) could be signed by any three of the nodes. In Fig. 3(a), nodes $1, 2$ and $3$ generate the partial signatures of message $m$ as follows: $PS(m, s1)$, $PS(m, s2)$ and $PS(m, s3)$. The three partial signatures could be combined to obtain the certificate $\mathcal{C}$, which is the same as it is signed by the CA's private key. In Fig. 3(b), nodes $4$ and $5$ are compromised by the attackers, but the two malicious nodes can not generate a valid certificate by themselves because at least three partial signatures are needed to be combined to sign a message.

This example shows that although a compromised node could also generate an incorrect partial signature, which would yield an invalid signature, a combiner can verify the validity of a computed signature using the CA's public key. In case the verification fails, the combiner tries another set of $t+1$ partial signatures. This process continues until the combiner constructs the correct signature from $t+1$ correct partial signatures.

The scheme described above was proposed in [62] for wireless ad hoc networks. Compared with ad hoc networks, WMNs are more favorable for utilizing the threshold cryptography key management scheme. First, WMNs is typically operator-managed, which makes it easier to pre-establish the distributed central authority (the

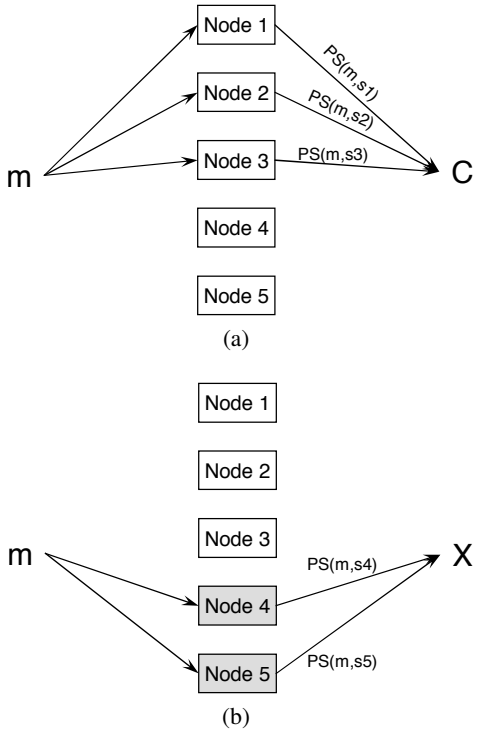



**Fig. 12.3.** (5, 3)-threshold signature.

CA's public key and private key shares) in WMNs than in ad hoc networks. Moreover, the nodes in WMNs are usually not mobile and hence do not rely on the battery power supply. Therefore, the asymmetric cryptography computation can be frequently processed in WMNs without much concern of the resource limitation.

There exists another public-key management scheme [41] in which two nodes can authenticate each other by finding a certificate chain between them. This scheme differs from the above in that it proposes a full-organized public key management system, where security does not rely on any trusted authority, not even in the initialization phase. Although the operator-managed WMNs do not require such a full self-organization key management, the certificate chain approach in [41] poses an interesting question: if $A$ can authenticate $B$ which in turn can authenticate $C$, is it 100% safe for $A$ to authenticate $C$? In other words, even if $A$ can authenticate $B$, should $A$ fully trust what $B$ trusts (that is, $C$ is authentic)? Furthermore, we can regard the whole process of authentication as a trust evaluation problem: "do I trust that you are who you claim you are?" The trust model for securing WMNs will be discussed in Section 12.4.5.

### 12.3.2 Secure Routing

In WMNs, the data travel via multiple wireless hops from the source node to its destination. The routing protocols for WMNs are designed to achieve:

- Self- configuration of the routing tables.
- Self-adaptation to changes in the wireless link quality.
- Maximized performance metrics such as end-to-end delay, throughput and packet loss rate.

The routing protocols for wireless ad hoc networks have also similar requirements such as routing through wireless multihop links, self-configuration and self-adaptation. Although very few routing protocols have been proposed specifically for WMNs, the similarities between WMNs and wireless ad hoc networks make it feasible for WMNs to borrow the ideas from the domain of wireless ad hoc networks, which have been extensively studied in the literature. For example, in 802.11s [63], the IEEE 802.11 standard for wireless LAN mesh networking, the Ad hoc On Demand Distance Vector (AODV) protocol [46] is extended to Radio Metric AODV (RM-AODV), an on demand routing protocol for wireless LAN mesh networks.

The self-configured and self-adapted wireless multihop routing mechanisms rely on the fact that all participating nodes cooperate with each other without disrupting the operation of the protocol. Without proper protection, the routing mechanisms could be attacked by both *external* and *internal* attacks [4].

#### External Attacks

Due to the shared nature of the wireless medium, anyone with a suitable hardware is able to send information into the medium. Indeed, external attackers can inject fabricated routing information into the network or maliciously alter the content of routing messages exchanged between the nodes. Therefore, the correctness of routing information exchange is vital to any routing protocols.

To secure routing, some proactive ad hoc routing protocols, such as DSDV [45] and OLSR [11], require that the routing messages are exchanged periodically between all the nodes so that each node has a view of the whole network's topology, based on which the routing decisions are made. The malicious routing messages with false topology information will make some nodes getting an incorrect view of the topology. On the other hand, for reactive routing protocols, such as AODV and DSR [26], the routing messages are exchanged between the source, destination and the intermediate nodes in order to find the best route after the source node initializing a route discovery process sends a packet to the destination. The route discovery process will then end up with a false route with the existence of the malicious routing massages.

To protect routing messages exchanging from attacks, the routing protocols need effective mechanisms to:

- Authenticate the received routing message to validate that it is sent by a legitimate node.

- Check the integrity of the received routing message to validate that it has not been altered by the attacker.

  Such mechanisms are often achieved by employing cryptographic solutions.

*Asymmetric Cryptography Approach*

As described in Section 12.3.1, asymmetric cryptography based authentication can prevent the fabricated routing information. When sending a routing message, the sender attaches the certificate signed by CA to the message and digitally signs the message with its private key. Upon receiving a routing message, the receiver first checks the validity of the certificate attached to the message using the CA's globally known public key and then checks the message's integrity using the digital signature and the sender's public key. ARAN [52] is an on demand protocol utilizing the digital signature scheme, in which the routing messages such as route discovery packet, reply packet and shortest path confirmation messages are signed by the sender and validated by the receiving node.

*Symmetric Cryptography Approach*

In this scheme, a single secret key is used for both encryption and decryption.

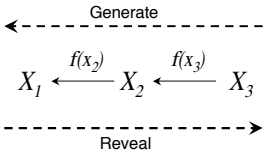



**Fig. 12.4.** One-way hash chain.

One of the most common schemes is one-way hash chains. As cited in [4], *"A one-way hash function is a function that takes an input of arbitrary length and returns an output of fixed length"*. Computing the input of a hash function from the output requires a huge amount of computation resource, so hash functions are computationally expensive to reverse. A hash chain is generated by applying a given hash function $f()$ repeatedly ($n$ times) to an initial input $x$ and obtaining a chain of outputs $f_i(x), i = 1, 2..., n$. The protocols utilizing one-way hash chains require that a shared secret, $f_j(x)$, exists so that the validity of $f_i(x)$ for $i < j$ can be checked by applying the hash function $j - i$ times on it and comparing the result with $f_j(x)$. Fig. 12.4 is an example of a one-way hash chain with length three. The chain is generated by applying the hash function $f()$ on $x_3$ such that $x_2 = f(x_3), x_1 = f(f(x_3)) = f_2(x_3)$. Since $x_1$ is a shared secret, $x_2$ and $x_3$ could be validated by checking if $f(x_2) = x_1$ and $f_2(x_3) = x_1$.

For instance, TESLA [47] is a broadcast authentication protocol based on one-way hash chains. In this protocol, the receivers need to buffer a message to wait for

the delayed key disclosure from the sender, which requires time synchronization. TESLA is employed by a distance vector protocol, SEAD [24], and a source routing protocol, Ariadne [23].

Compared with the digital signature scheme, the one-way hash chain scheme has the advantage of light weight computation cost and no need to maintain a globally trusted CA, but it requires clock synchronization of all the nodes in network. Furthermore, in TESLA, a received message has to be stored in buffer waiting for the disclosed key to authenticate it before being processed, which degrades the performance of the network.

To overcome such limitations, a hybrid approach, SAODV [60] was proposed where the non-mutable fields of the routing messages are signed by asymmetric cryptography while the mutable field, hop count, is authenticated using a hash chain so that the expensive asymmetric cryptographic computation is only needed for the source and destination nodes and the intermediate nodes authenticate the hop count using hash function.

The one-way hash chain scheme is more favorable for wireless ad hoc networks in which the nodes are battery-powered and the computation resource is limited. Furthermore, the node mobility in wireless ad hoc networks makes it difficult to maintain an online CA available for all the nodes. However, the nodes in WMNs are not mobile and they do not rely on battery power supply. So the digital signature scheme is a better choice for WMNs if the clock synchronization is hard to achieve.

### Internal Attacks

If an attacker gains full control of a legitimate node, the cryptographic approaches will not be able to prevent the attacks launched from the node because the node has valid cryptographic keys and the messages sent by the node are also cryptographically valid. The compromised nodes could attack the routing mechanisms by generating false routing information, scheduling the data packets forwarding for their own benefits, selectively forwarding the packets, or not forwarding any packet at all. Here, we discuss some countermeasures to internal attacks.

#### *Packet Leash*

In [25], a challenging attack, called the wormhole attack was defined. If an attacker gets control of two nodes with a wired communication link (tunnel) between them, the wormhole attacks could be launched by sending all the packets received from one node through the tunnel and replaying these packets at the other end of the tunnel. Fig. 12.5 shows an example of wormhole attack [4]. Because the packets through the tunneled link ($A \rightarrow B$) arrive sooner than the packets through the multihop wireless links ($1 \rightarrow 2 \rightarrow 3 \rightarrow 4$), nodes 2 and 3 are excluded from the network, and the traffic between nodes 1 and 4 is completely under the control of the attacker.

The Packet Leash solution [25] is to add some extra information to each message at the sender side in order to allow the receiver to determine if the packet has traversed an unrealistic distance. The extra information could be a precise timestamp,

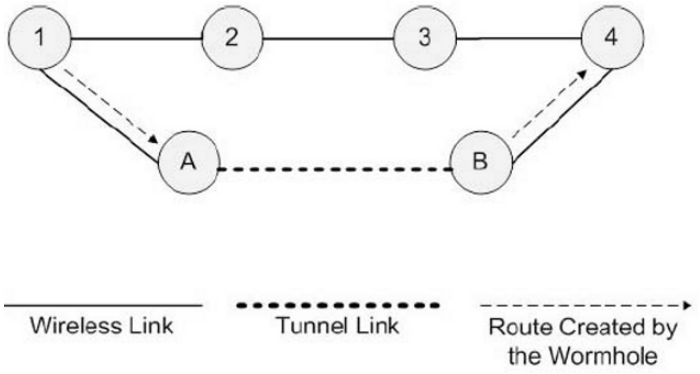



**Fig. 12.5.** A wormhole attack performed by colluding malicious nodes A and B.

which requires extremely precise clock synchronization, or the location information with a timestamp, which requires less precise clock synchronization.

*Neighbor Monitoring*

The neighbor monitoring approach to discover misbehaving nodes takes the advantage of the broadcast nature of wireless network: any packet sent in to the air can be overheard by the neighbor nodes. After a node sends a packet to its neighbor, it could monitor the behavior of its neighbor to see whether it forwards the packet to the next hop without any misbehavior. Each node maintains a rating record of all the nodes it knows, and the misbehaviors of a particular node being detected cause the rating to decrease. The low rating nodes are considered misbehaving or non-trust nodes so that they will not be included in the route of forwarding packets from source to the destination nodes.

Based on this approach, two other solutions [42] and [8] were proposed to defend against packet forwarding attacks.

*Byzantine Failure Resilience*

In [5], an on-demand secure routing protocol was proposed that is resilient to Byzantine failures caused by Byzantine behavior, which is defined as "any action by an authenticated node that results in disruption or degradation of the routing service". The failure refers to "any disruption that causes significant loss or delay in the network". The detection of such failures is based on acknowledgements (*acks*). The destination node sends an *ack* back to the source node when receiving a packet. If an *ack* is not received after a certain time, the source node assumes it has been lost. The number of lost (to the same destination) exceeding a threshold triggers the Byzantine fault detection.

Fig. 12.6 illustrates the detection process [5]. The source node launches a binary search of all the links along the path by probing the intermediate nodes. The normally behaving nodes send *acks* back to the source when receiving the probe. Half of the

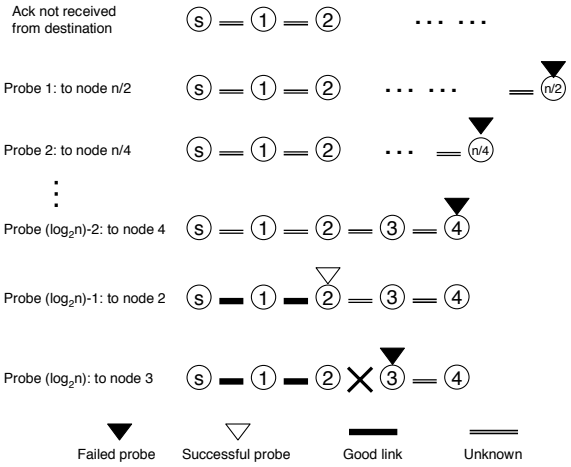



**Fig. 12.6.** Byzantine fault detection: The faulty link is located after $\log n$ probes.

links are excluded from the suspects of failure for each probe. The faulty link will be identified after $\log n$ probes, where $n$ is the number of hops between the source and destination. After the failure is located, the source node will start a new route discovery process and try to bypass the faulty link.

Fig. 12.7 summarizes various schemes for secure routing that may be applied to WMNs.

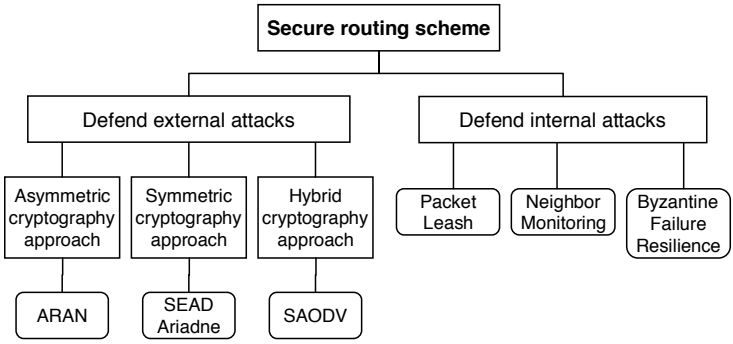



**Fig. 12.7.** Secure routing protocols.

### 12.3.3 Secure Location Information

As mentioned before, most routing protocols in WMNs are adopted from ad hoc networks, including both topology-based and geographic routing schemes. For geographic routing schemes [6, 12, 19, 22], the location information of the mesh routers are crucial to multihop routing schemes and thus subject to attack.

For securing location information, two general methods are currently employed: correctly compute location information, and verify location claims.

Generally speaking, a mesh router's location can be determined either with the help of GPS or some location-known beacons. The goal of the first approach is to ensure the accuracy of location computation even when the calculation is under attack. For example, although GPS is the most common approach to get the geographic position information, no secure protection for public civilian GPS makes it vulnerable to different kinds of attacks [34]. As an example, a *signal-synthesis* attack can fool a receiver to connect to a device present at some pretended location. Similarly, *selective-delay* attack can convert a signal received at time $t$ and position $r$ into another signal that would have been received at earlier time $t'$ and position $r'$.

To defend such attacks, an information-hiding based asymmetric security mechanism was proposed in [34]. The essence of the scheme is to introduce time asymmetry through a delayed disclosure of despreading key. Specifically, when a spread-spectrum broadcast signal temporarily hidden in the background noise is transmitted, the receivers store the whole radio band in buffer. And the despreading key is not published until the delay is larger than the uncertainty of the local clock in the receiver. In this way, both signal-synthesis and selective-delay attack can be easily detected.

For the schemes that utilize beacons, a cryptographic-based scheme, SeRLoc (Secure Range-independent Location), was proposed to enable the nodes to determine their location even in the presence of malicious adversaries [37]. In SeRLoc, some nodes which are equipped with directional antennas and have acquired the location and the orientation through GPS receivers are termed "locator". Each locator transmits different beacons at each antenna sector containing its coordinates and the angles of the antenna boundary lines with respect to a common global axis. The nodes will collect the beacons from all locators they can hear and then determine their location. To protect the localization information, a global symmetric key is shared between nodes and locators. Moreover, every sensor shares a symmetric pairwise key with every locator so that the beacons from each locator can be authenticated. The analysis shows SeRLoc is robust against several attacks including wormhole attack, Sybil attack and compromised nodes. However, one limitation of this scheme lies in that it assumes locators are always trusted and cannot be compromised by an adversary.

Besides directly securing the location calculation, the location information may also be verified from spoofing. Due to the fact that the mesh routers in WMNs are usually static, a claim of location information made by the mesh routers to the mesh clients would often be more than just location calculation. Therefore, location verification would be more agreeable in WMNs. To validate a node is in a region of its po-

sition claim, different techniques such as exampling by public key based challenge-response protocol [7] and robust statistical methods [36, 39, 40] can be employed. In addition, by exploiting the physical properties of sound and RF signal propagation in wireless communication, a simple protocol called "Echo", was proposed [53] that requires no cryptography nor time synchronization. Another mechanism, called Verifiable Multilateration (VM), achieves both secure position computation and location verification [9]. All such schemes proposed for wireless sensor networks seem to be applicable to WMNs as well.

### 12.3.4 Modeling Virus Propagation

Given the fast emergence of computer viruses in the host computers and the Internet, the threat of virus in wireless networks is not an unrealistic panic. In fact, there have been some viruses that spread over the air, such as the Brador virus [55] and the Cabir worm [18] for the mobile devices, the evil twin and the promiscuous client for Wi-Fi users [16].

In order to effectively defend the virus attack, one important issue is how to model the virus propagation to get a better understanding on the virus behavior in wireless networks. Inspired by biological modeling techniques, some researchers have adopted Epidemic theory to model the virus propagation problem.

Epidemic theory [3] is the study of the dynamics of how contagious diseases spread in a population, resulting in an epidemic. It can mathematically model the progress of the infectious diseases and measure its outcome in relation to a population at risk. In general, the population is divided into three groups: the *susceptible* (S), who are healthy and are subjective to catching the disease; the *infected* (I), who have the disease and can transmit it; and the *removed* (R), who have had the disease and are recovered now. In general, there are two popular models to characterize the infection spread: *Susceptible Infected Susceptible* (SIS) and *Susceptible Infected Recovered* (SIR). The difference between these two models is the following. For an individual who acquires infection, this individual can becomes susceptible again after some infectious period in the former model, while in the latter model, the individual becomes immune to further infections after recovery.

An important aspect in Epidemic theory is that the phase transition of the spreading process is dependent on an threshold of the epidemic parameter. That is, when the epidemic parameter is above the threshold, the infection will spread out and become persistent; on the contrary, if the parameter is below the threshold, the infection will die out. Therefore, identifying this threshold value is critical in the study of how an epidemic spreads and how it can be controlled [13].

Epidemic theory has been employed to investigate virus spreading problem not only in wirelined networks, but also in wireless networks. Here, we list two schemes that apply Epidemic theory to model the worm and compromised nodes propagation, respectively.

*Topologically-Aware Worm Propagation Model (TWPM)*

TWPM was proposed in [33] for wireless sensor networks. By parameterizing the effects of physical, MAC, and network layers on the worm propagation, the authors incorporated all these parameters in the SIS model and analytically derived the worm propagation model from a partial differential equation. With some assumptions including regular two-dimensional grid topology and constant infection rate, they also obtained a closed-form expression for the TWPM model.

Although this work is originally proposed for wireless sensor networks, it has the potential to adapt to WMNs by taking real topology into account. In WMNs, most mesh routers have a neighbor list, either for routing purpose or infrastructure maintenance. Unfortunately, the scanning worms, called *topologically-aware worms*, can take advantage of this list and spread the infection quite effectively by just communicating to its next-hop neighbors.

*Modeling Node Compromise Spread*

Unlike TWPM using a differential equation approach to solve the problem, a network and graph theory based technique was proposed to model node compromise spread in wireless sensor networks [14].

In general, no matter whether its is a sensor network or a mesh network, for secure communication, a secret key used to encrypt the messages is shard between each communication party. However, without physical protection, the nodes are subject to capture. Once a node is captured, its keys are known by the attackers, thus affecting communications with all the compromised node involved. The work in [14] studied how an adversary capturing one or two nodes and thereby extracting the secret keys can possibly propagate the node compromise to the whole network.

By constructing a random graph model of the key sharing overlay graph of the sensor network and presenting the compromised propagation model as a Poisson process, this work investigated the probability of a breakout (when the whole network is compromised) and also computed the size of the compromised clusters of nodes under no breakout. Additionally, the effects of two scenarios – recovery and no recovery – on the compromised nodes recovery were analyzed in this work.

Although this scheme is proposed for wireless sensor networks, the essential idea and basic assumptions are still valid for WMNs. Therefore, they shed some light on modeling virus propagation in WMNs.

## 12.4 Summary and Open Problems

Wireless mesh networks (WMNs) possess some nice features and promise to offer better wireless network connectivity and larger coverage area. On the other hand, these features also pose significant challenges to the network security. In this chapter, we have reviewed some existing solutions that could potentially be employed to secure WMNs. The threshold signature and TESLA schemes could be utilized to

realize the authentication between mesh routers. Such authentication schemes also help routing protocols such as ARAN, SEAD and Ariande to defend against external attacks. The Packet Leash, neighbors monitoring and Byzantine failure resilience solutions provide possible approaches to detect and countermeasure internal attacks to the wireless multihop routing protocols. The secure location information solutions for wireless sensor networks can help securing the geographic routing schemes. However, there are still a lot of open problems for security in WMNs that need further investigation. These are discussed below.

### 12.4.1 Secure Medium Access Control

The IEEE 802.11 medium access control (MAC) protocol has been adopted as the de facto MAC scheme of WMNs in many research projects and commercial products ( [10,44,51]). The cryptographic approach for securing the 802.11 MAC protocol had evolved from Wired Equivalent Privacy (WEP) protocol to IEEE 802.11i standard [64].

In IEEE 802.11i, an authentication server (AS) is incorporated to authenticate the mobile node (MN) that tries to associate with the mesh access point (AP). IEEE 802.11s [63], the IEEE standard for wireless LAN mesh networking, utilizes IEEE 802.11i based security mechanism to provide link-by-link security in WLAN mesh networks. According to IEEE 802.11s, the AS is collocated with a mesh point (MP) or located in a remote entity to which an MP has a secure connection. It is assumed by the standard that an MP could establish a secure connection to the remote AS after establishing a secure connection with the MP that collocates with the AS or has a secure connection with the AS. But the standard neither proposes how to establish the secure connection nor evaluates the practicality of establishing such a secure connection. Furthermore, the 802.11i security framework has a centralized structure in which the MNs submit authentication request to AP which in turn communicates with AS to decide whether to authenticate or not. But such a hierarchical structure is not suitable for WMNs in which two MPs need to authenticate each other before they start communication; in other words, the MPs are both the authentication supplicants and the authenticator. This mutual authentication requires that both MPs have a secure connection with AS, which is impractical in WMNs.

The cryptographic approach to secure the MAC protocol is to answer the question “who can utilize the mesh medium?” However, it does not address the issue of fair utilization of the medium. The attacks to the backoff scheme of IEEE 802.11 MAC protocol break the fairness of using the wireless medium. The IEEE 802.11 MAC protocol uses CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance) scheme to reduce the probability of collisions in accessing the medium. If the sender senses the channel to be busy before transmission, it defers the transmission for a random backoff time. Simulation results reported in [35] show that if a misbehaving node selects smaller backoff time than other nodes complying the protocol, it will obtain more than its fair share of the bandwidth and degrade the throughput of well-behaved nodes.

A modification to IEEE 802.11 protocol was proposed in [35] to countermeasure this attack. In this proposal, the sender does not decide the value of random backoff time. Instead, the receiver selects a backoff time value and sends it to the sender. The receiver can identify whether a sender deviated the protocol by monitoring the time intervals between the sender's transmissions and comparing them with the backoff time assigned to the sender. If the deviation is identified, the receiver will penalize the sender by assigning larger backoff values to it than those assigned to normal nodes. Such a scheme could restrict the selfish nodes to get more bandwidth share, but it could do nothing to prevent the malicious nodes from attacking the backoff scheme if the malicious nodes that do not care how much bandwidth to share, keep transmitting data without backoff time at all. Such an attack is a denial of service (DoS) attack to the MAC layer protocol, which is still an open issue for the wireless networks relying on the CSMA/CA scheme.

### 12.4.2 Defense Against DoS Attacks

Denial of Service (DoS) attacks can reduce the availability of resource and result in massive service disruption. A robust WMN application should be resilient to DoS attacks and be able to defend against such attacks launched either by the end devices or other adversaries.

DoS attacks could happen at all the layers in the protocol stack from the physical to the application layer [56]. Usually different approaches are employed in different layers of the protocol. For instance, at the physical layer, the most common defense against DoS (e.g., jamming) is spread spectrum. At the MAC layer, some special measures such as rate limitation and error correcting code may be used to defend against DoS attacks. Although most of routing schemes in WMNs are adopted from ad hoc networks, the characteristics of WMNs, especially multihop routing, should be taken into account for the defense mechanisms at the network layer. In particular, a poorly designed multihop routing scheme may introduce traffic unfairness or starvation, which even leads to DoS for some mesh routers that are close to the backbone network.

Even though there is no universal way to defend against DoS attacks, a systematic framework that can comprehensively consider all these issues at the beginning design phase would be more effective. Furthermore, an integrated, cross-layer security solution is more desirable.

### 12.4.3 Embedded Security Schemes vs. System Level Monitoring

The majority of the current security mechanisms are embedded in the network protocols, so they usually focus on some particular attacks at a specific layer and are efficient for outside (external) attacks.

An alternate approach is to design a cross-layer framework that can monitor in real time the whole network to detect attacks and respond promptly. Compared with the embedded schemes, the monitoring framework can work as an intrusion detection system (IDS) to detect any real-time abnormality. Since the intrusion includes not

only the attacks launched by the outsiders but also the misuse from the inside, it is more effective and flexible to defend insider (internal) attacks.

However, WMNs pose new challenges for designing intrusion detection schemes. First, mesh routers that are usually not physically protected are subject to capture. Once a mesh router gets captured, all of its secret information including keys is disclosed to the adversary. These corrupted mesh routers not only compromise the whole network security, but can also modify the network configuration or inject false information to disturb the routing schemes. Moreover, the delay introduced by multi-hop communication causes difficulty for traffic monitoring. Therefore, how to detect the corrupted mesh routers and inform the whole network in a timely manner is still an open problem in WMNs.

### 12.4.4 Integration Issues

One main advantage of WMNs is that it enables us to integrate various existing networks such as Wi-Fi, cellular networks, sensor networks, etc, through the gateways. However, this benefit also brings related vulnerability in WMNs.

Various (heterogeneous) networks as part of WMN clients imply their properties may have significant differences as well. For example, although the public key cryptography is a common approach for most networks for authentication, it may be computationally too costly for sensor networks. Naturally, WMNs should be able to customize the security schemes according to the characteristics of network clients while not compromising the security features of the overall network. Therefore, the interworking of several different types of networks poses a new challenge to securing WMNs.

### 12.4.5 Trust Relationships

Some security issues, such as establishing certificate chains [41] and collaborating between mesh routers to implement routing protocols or reduce the authentication delay by sharing the security key [21], imply the existence of trust relationships between different entities of WMNs. A mechanism to define how to establish and quantify such trust relationships could help the mesh routers to make proper decisions in the presence of potential attacks, and thus improve the reliability and robustness of the networks.

Although the trust and reputation systems have been extensively and successfully applied in E-commerce applications [50, 57, 59], public key authentication [20, 38, 43, 49], peer-to-peer networks [32, 58], mobile ad hoc networks [17, 54] and wireless sensor networks [61], some characteristics of WMNs differ from these networks or applications. Therefore, there is a need for new trust establishment and management in WMNs.

On one hand, the existence of infrastructure/backbone in WMNs favors the trust establishment. Unlike ad hoc networks supporting node mobility, most of the mesh routers are static. Intuitively, a trust infrastructure could be established on top of the

fixed WMN backbone. And once it is created, the entities and their trust chains will be permanently present.

On the other hand, the erroneous and uncertain nature of wireless links cannot guarantee stable connectivity between the mesh routers even when they physically exist. Moreover, the quality of wireless links may also change frequently. While in traditional applications, the trust relation are usually established and updated based on a long time observation, in WMNs, however, the trust infrastructure should have the capability to respond to temporary disconnections between the mesh routers. Hence, the evaluation of the trust relation must be fast enough to promptly reflect the instantaneous changes in the environment.

Moreover, the mesh routers in WMNs can only monitor their neighbors within the radio range which means that the trust relation is establishment locally and thus distributed. Therefore, when the packets are routed via multihop links, the trust should be uniformly evaluated. As a result, it is necessary to develop some scheme that can manage the trust propagation. In [31], a method of Trust Network Analysis with Subjective Logic (TNA-SL) was designed based on graph simplification and trust derivation with subjective logic. TNA-SL expresses and propagates both positive and negative trust values and takes the confidence level (certainty) into account, which might make it an appropriate model for calculating and evaluating quantified trust in WMNs. There exist other works that were proposed to define the trust transitivity and manage trust propagation [27–30]. However, the validation and performance of such schemes under real WMNs applications needs further investigation.

Besides the above issues, a node in a WMN may not always be able to monitor its neighbors if they are using different radio interfaces. In addition, how to effectively make the trust relation globally available to the whole network is also a challenging problem in WMNs. However, the trust based system fits the characteristics of WMNs and provides a promising solution to secure WMNs.

## Conclusion

In this chapter, we have addressed some security issues in WMNs and surveyed state-of-the-art solutions that have either been applied to WMNs or have the potential to be adopted. It is worth noting that no panacea exists that can solve all the problems identified. In fact, there are currently more open problems that need further investigation than solutions to secure WMNs. Depending on the specific applications and requirements, some approaches need to work together to achieve the desired security, or a cross-layer solution should be developed while designing WMN applications.

## References

1. I. Akyildiz and W. Wang, “A survey on wireless mesh networks,” *IEEE Communications Magazine*, vol. 43, no. 9, pp. S23-S30, Sept. 2005.

2. M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. ACM MobiCom'05*, Cologne, Germany, 58–72 Aug, 2005.
3. R. M. Anderson and R. M. May, *Infectious Diseases of Human: Dynamics and Control*, Oxford Univ. Press, Oxford, 1991.
4. P. G. Argyroudis and D. O'Mahony, "Secure routing for mobile ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 7, no. 3, pp. 2-21, Atlanta, GA, USA, 2005.
5. B. Awerbuch, D. Holmer, C. Nita-Rotaru, and H. Rubens, "An on-demand secure routing protocol resilient to byzantine failures," in *Proc. ACM WiSe'02*, pp. 21-30, Sept. 2002.
6. P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks," *Wireless Networks*, vol. 7, no. 6, pp. 609-616, Kluwer Academic Publishers, 2001.
7. S. Brands and D. Chaum, "Distance-bounding protocols," *EUROCRYPT '93: Workshop on the Theory and Application of Cryptographic Techniques on Advances in Cryptology*, vol. 765 of LNCS.
8. S. Buchegger and J. L. Boudec "Performance analysis of the CONFIDANT protocol," in *Proc. ACM MobiHoc'02*, pp. 226-236, Lausanne, Switzerland, Jun. 2002.
9. S. Capkum and J. Hubaux, "Secure positioning of wireless devices with application to sensor networks," in *Proc. IEEE INFOCOM'05*, pp. 1917-1928, Mar. 2005.
10. Overview: Wireless Mesh Networking, http://www.cisco.com/en/US/netsol/ns175/networking_solutions_products_generic_content0900aecd80529a46.html, 2007.
11. T. H. Clausen, G. Hansen, L. Christensen, and G. Behrmann, "The optimized link state routing protocol, evaluation through experiments and simulation," in *4th International Symposium on Wireless Personal Multimedia Communications*, Aalborg, Denmark, 2001.
12. S. Datta, I. Stojmenovic, and J. Wu, "Internal node and shortcut based routing with guaranteed delivery in wireless networks," in *Proc. IEEE International Conference on Distributed Computing and Systems Workshops, Cluster Computing*, pp. 461-466, April 2001.
13. P. De and S. K. Das, "Epidemic models, algorithms and protocols in wireless sensor and ad hoc networks," *Handbook on Wireless Sensor Networks*, John Wiley, 2007.
14. P. De, Y. Liu, and S. K. Das, "Modeling node compromise spread in sensor networks using Epidemic theory," in *Proc. IEEE WOWMOM'06*, pp. 237-243, Washington, DC, Jun. 2006.
15. Y. Desmedt, "Threshold cryptography," *European Transactions on Telecommunication*, vol. 5, no. 4, pp. 449-457, 1994.
16. C. Elliott, *wireless threats to your business,* http://www.microsoft.com/smallbusiness/resources/technology/broadband_mobility/6_wireless_threats_to_your_business.mspx.
17. L. Eschenauer, V. Gligor, and J. Baras "On trust establishment in mobile ad-hoc networks," in *Proc. of 10th International Workshop on Security Protocols*, Cambridge, UK, April 2002.
18. P. Ferrie, P. Szor, R. Stanev, and R. Mouritzen, "Security response: SymbOS.Cabir", 2004. Symantec Corporation.
19. H. Frey, "Scalable geographic routing algorithms for wireless a hoc networks," *IEEE Network Magazine*, Jul./Aug. pp. 18-22, 2004.
20. R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proc. of International World Wide Web Conference*, pp. 403-412, May 2004.

21. J. Hassan, H. Sirisena, and B. Landfeldt, "Trust-based fast authentication in multiowner wireless networks," *IEEE Transactions on Mobile Computing*, in press.
22. M. Heissenbuttel and T. Braun, "BLR: Beacon-less routing algorithm for mobile ad hoc networks," *Computer Communications Journal*, vol. 27, no. 11, pp. 1076-1086, July 2004.
23. Y. C. Hu, D. B. Johnson, and A. Perrig, "Ariadne: A secure on-demand routing protocol for ad hoc networks," in *Proc. ACM MobiCom'02*, pp. 12-23, Sept. 2002.
24. Y. C. Hu, D. B. Johnson, and A. Perrig, "SEAD: Secure efficient distance vector routing for mobile wireless ad hoc networks," in *Proc. WMCSA'02*, pp. 3-13, Jun. 2002.
25. Y. C. Hu, D. B. Johnson, and A. Perrig, "Packet leashes: A defense against wormhole attacks in wireless networks," in *Proc. IEEE INFOCOM'03*, vol. 3, pp. 1976-1986, March/April 2003.
26. D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, Kluwer Academic Publishers, 1996.
27. A. Josang, "Trust-based decision making for electronic transactions," L. Yngstrom and T. Svensson, editors, in *Proc. NORDSEC'99*. Stockholm University, Sweden, Stockholm University Report 99-005, 1999.
28. A. Josang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 279-311, June 2001.
29. A. Josang and R. Ismail, "The Beta reputation system," in *Proc. 15th Bled Conference on Electronic Commerce*, Bled, Slovenia, June 2002.
30. A. Josang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618-644, March 2007.
31. A. Josang, E. Gray, and M. Kinateder, "Simplification and analysis of transitive trust networks," *Web Intelligence and Agent Systems Journal*, vol. 4, no. 2, pp. 139-161, 2006.
32. S. Kamvar, M. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in P2P networks," in *Proc. WWW'03*, pp. 640-651, Budapest, Hungary, May 2003.
33. S. A. Khayam and H. Radha, "A topologically-aware worm propagation model for wireless sensor networks," in *Proc. ICDCSW'05*, pp. 210-216, Washington, DC, USA, June 2005.
34. M. Kuhn, "An asymmetric security mechanism for navigation signals," in *Proc. of the Information Hiding Workshop*, pp. 23-25, May 2004.
35. P. Kyasanur and N. H. Vaidya, "Detection and handling of MAC layer misbehavior in wireless networks," in *Proc. International Conference on Dependable Systems and Networks*, pp. 173-182, 2003.
36. L. Lazos, R. Poovendran, and S. Capkum, "ROPE: Robust position estimation in wireless sensor networks," in *Proc. IEEE IPSN'05*, pp. 324-331, April 2005.
37. L. Lazos and R. Poovendran, "SeRLoc: Secure range-independent localization for wireless sensor networks," in *Proc. ACM WiSe'04*, pp. 21-30, Philadelphia, PA, USA, Oct. 2004.
38. R. Levien and A. Aiken, "Attack-resistant trust metrics for public key certification," in *Proc. of 7th USENIX Security Symposium.*, pp. 229-242, Jan. 1998.
39. Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," in *Proc. IEEE IPSN'05*, pp. 91-98, April 2005.
40. D. Liu, P. Ning, and W. Du, "Attack-resistant location estimation in sensor networks," in *Proc. IEEE IPSN'05*, pp. 99-106, April 2005.
41. H. Luo, P. Zerfos, J. Kong, S. Lu, and L. Zhang, "Self-securing ad hoc wireless networks," in *Proc. IEEE ISCC'02*, pp. 567-574, July 2002.

42. S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," in *Proc. ACM MobiCom'00*, pp. 255-265, Boston, Massachusetts, USA, August 2000.
43. U. Maurer, "Modeling a public-key infrastructure," in *Proc. Eur. Symp. Res. Comput. Security*, vol. 1146, pp. 325-350, Lecture Notes in Computer Science, 1996.
44. Motorola. Motorola's Mesh Networking Technology & Industry (IEEE) Standards, http://www.motorola.com/mesh/pages/technology/industry_standards.htm.
45. C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proc. ACM SIGCOMM'94* , pp. 234-244, London, UK, 1994.
46. C. Perkins, "Ad hoc On-Demand Distance Vector (AODV) routing," *IETF RFC 3561*, 2003.
47. A. Perrig, R. Canetti, D. Tygar, and D. Song, "The TESLA broadcast authentication protocol," *RSA Cryptobytes*, vol.5, no. 2, pp. 2-13, 2002.
48. A. Raniwala and C. Tzi-cker, "Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network," in *Proc. IEEE INFOCOM'05*, pp. 2223-2234, March 2005.
49. M. K. Reiter and S. G. Stubblebine, "Resilient authentication using path independence," *IEEE Transactions on Computers*, vol. 47, no. 12, pp. 1351-1362, Dec. 1998.
50. P. Resnick and R. Zeckhauser, "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," *The Economics of the Internet and E-Commerce*, M. R. Baye, editor, volume 11 of Advance in Applied Microeconomics, Amsterdam, Elsevier Science, 2002.
51. Roofnet. http://pdos.csail.mit.edu/roofnet.
52. K. Sanzgiri, B. Dahill, B. Levine, C. Shields, and E. M. BeldingRoyer, "A secure routing protocol for ad hoc networks," in *Proc. IEEE ICNP'02*, pp. 78-87, Nov. 2002.
53. N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in *Proc. ACM Workshop on Wireless Security*, pp. 1-10, San Diego, CA, 2003.
54. Y. Sun, W. Yu, Z. Han, and K. J. R. Liu, "Information theoretic framework of trust modeling and evaluation for ad hoc networks," *IEEE JSAC*, Special Issue on Security in Wireless Ad hoc Networks, vol. 24, no. 2, pp. 305- 317, Feb. 2006.
55. R. Wong and I. Yap, "Security information: Virus encyclopedia: Technical details," *Trend Micro Incorporated*, 2004.
56. A. Wood and J. Stankovic, "Denial of service in sensor networks," *IEEE Computer*, vol. 35, no. 10, pp. 54-62, 2002.
57. B. Yu and M. P Singh, "A social mechanism of reputation management in electronic communities," in *Proc. of the 4th International Workshop on Cooperative Information Agents*, pp. 154-165, July 2000.
58. B. Yu, M. Singh, and K. Sycara, "Developing trust in large-scale peer-to-peer systems," in *Proc. of 1st IEEE Symposium on Multi-Agent Security and Survivability*, pp. 1-10, Aug. 2004.
59. G. Zacharia, A. Moukas, and P. Maes, "Collaborative reputation mechanisms in electronic marketplaces," in *Proc. IEEE HICSS'99*, pp. 8026, Jan. 1999.
60. M. G. Zapata and N. Asokan, "Securing ad hoc routing protocols," in *Proc. IEEE WiSe'02*, pp. 1-10, Atlanta, GA, USA, 2002.
61. W. Zhang, S. Das, and Y. Liu "A trust based framework for secure data aggregation in wireless sensor networks," in *Proc. IEEE SECON'06*, pp. 60-69, Sept. 2006.
62. L. Zhou and Z. J. Haas, "Securing ad hoc networks," *IEEE Network*, vol. 13, no. 6, pp. 24-30, 1999.

63. "Joint SEE-Mesh/Wi-Mesh Proposal to 802.11 TGs, 2006".
64. "IEEE Std 802.11i/D4.1", *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Security Enhancements*, 2003.

# Index