وارسی صحت پرسمان روی پایگاه داده برون‌سپاری شده

محمدعلی هادوی

مجتمع دانشگاهی فناوری اطلاعات، ارتباطات و امنیت – دانشگاه صنعتی مالک اشتر

hadavi@mut.ac.ir

چشم انداز

- مقدمه
  - نیازمندی‌های درستی پاسخ
  - رویکرد کلی برای وارسی پاسخ
- رویکردهای وارسی درستی پاسخ پرسمان
  - ساختارهای داده احراز اصالت
    - درخت چکیده‌ساز مرکل
    - زنجیره امضا
  - روش‌های مبتنی بر افزونگی داده
- جمع‌بندی و چالش‌های پژوهشی

مقدمه

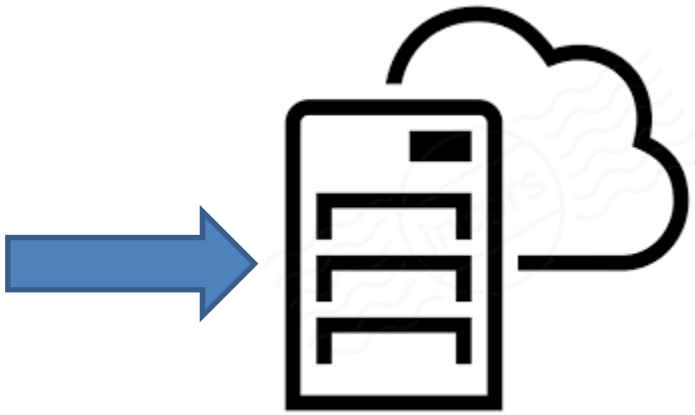
کاهش سطح اعتماد به کارپذیر ابر

- افزایش توان مهاجم
  - مهاجم فعال (خراب‌کار)
- کارپذیر ابری خراب‌کار؟!
  - فقدان داده و پنهان‌کاری ابر (حفظ اعتبار)
  - حملات به ابر و کنترل کارپذیرهای ابری توسط حمله‌کننده
  - خرابکاری‌های ناشی از بدافزارها
  - ...

درستی پاسخ پرسمان

مالک داده

Name	Salary
A	$5000
B	$3500
C	$9000
D	$6000



کارپذیر ابر

```
SELECT Name
FROM  Employees
WHERE Salary > 7000
```

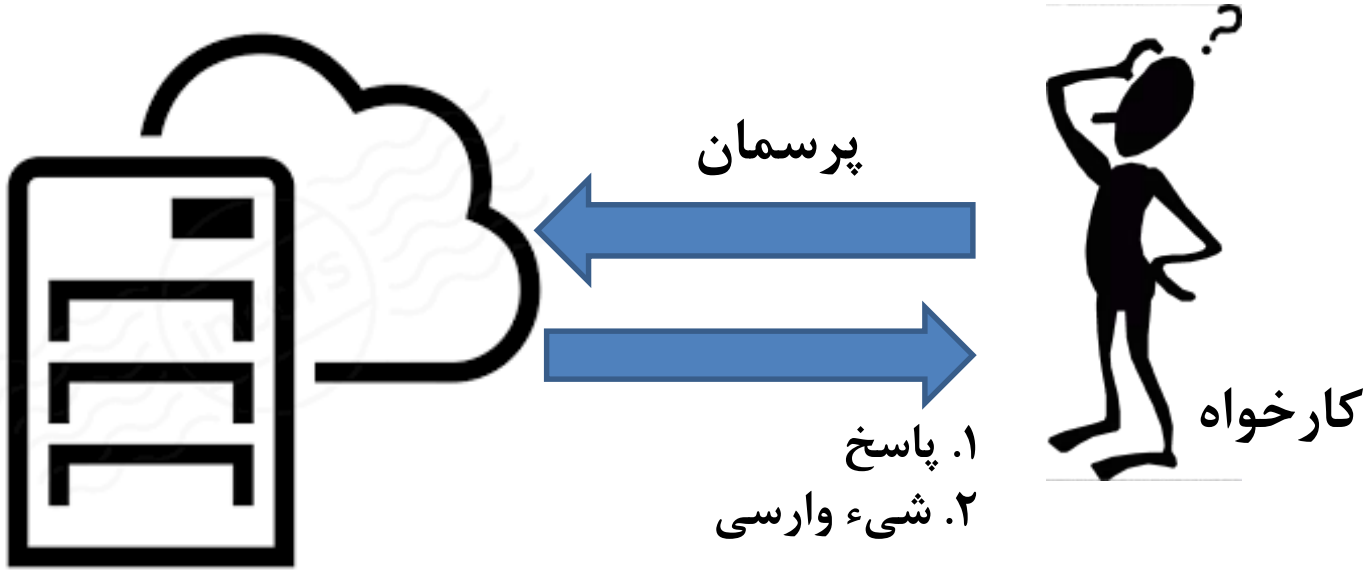
کاربر

R

آیا R یک پاسخ درست است؟

پاسخ درست؟

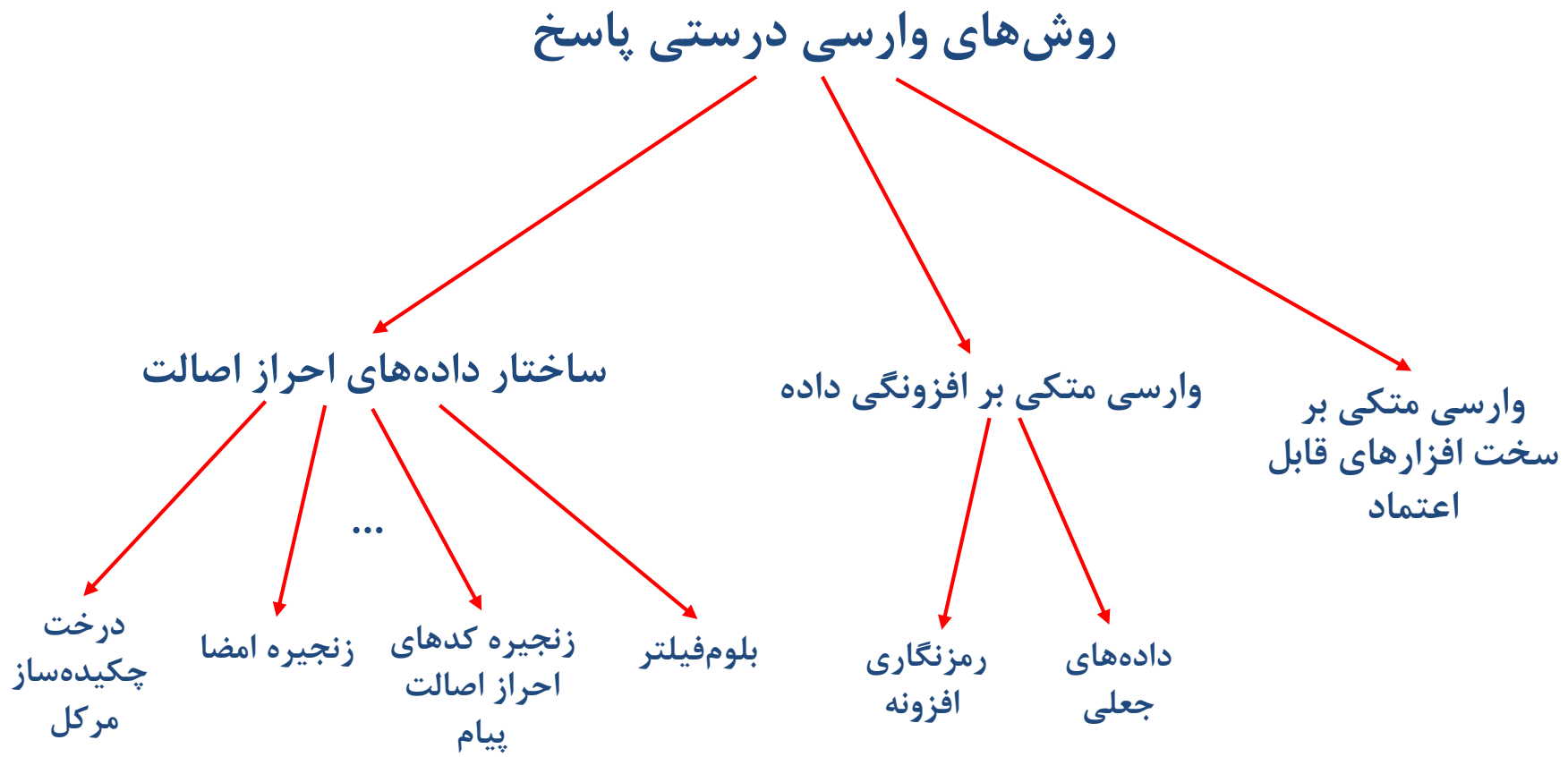
- پاسخ پرسمان درست است اگر
  - صحیح و اصیل باشد
  - کامل باشد
  - به‌هنگام (تازه) باشد
  - محیط‌های پویا



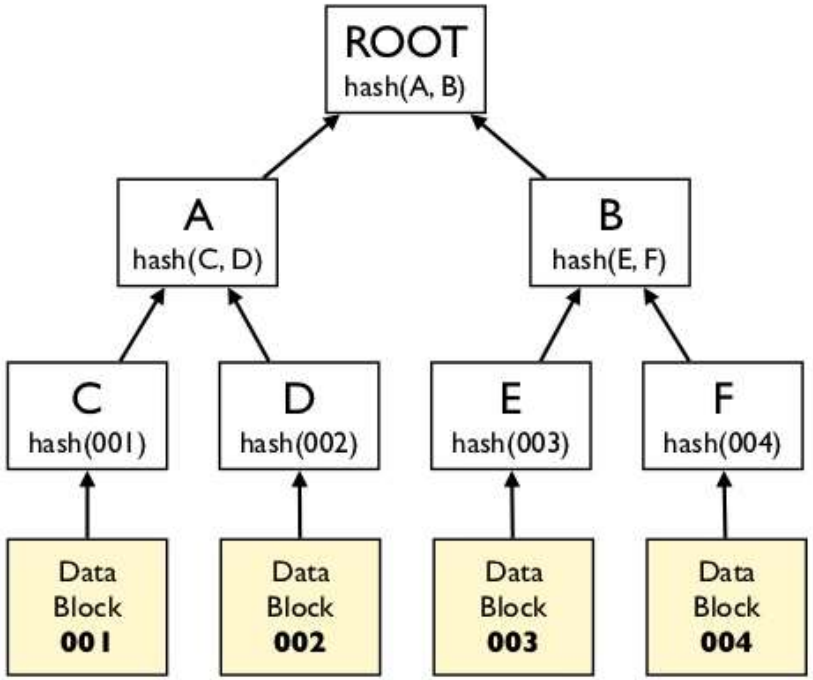
- ایده‌ی اصلی
  - وارسی پاسخ متکی بر ارسال اطلاعات وارسی (شیء وارسی) همراه با پاسخ پرسمان
    - تضمین درستی پاسخ
    - احتمال درستی پاسخ

رویکردهای وارسی درستی پاسخ پرسمان

رویکردهای موجود در یک نگاه



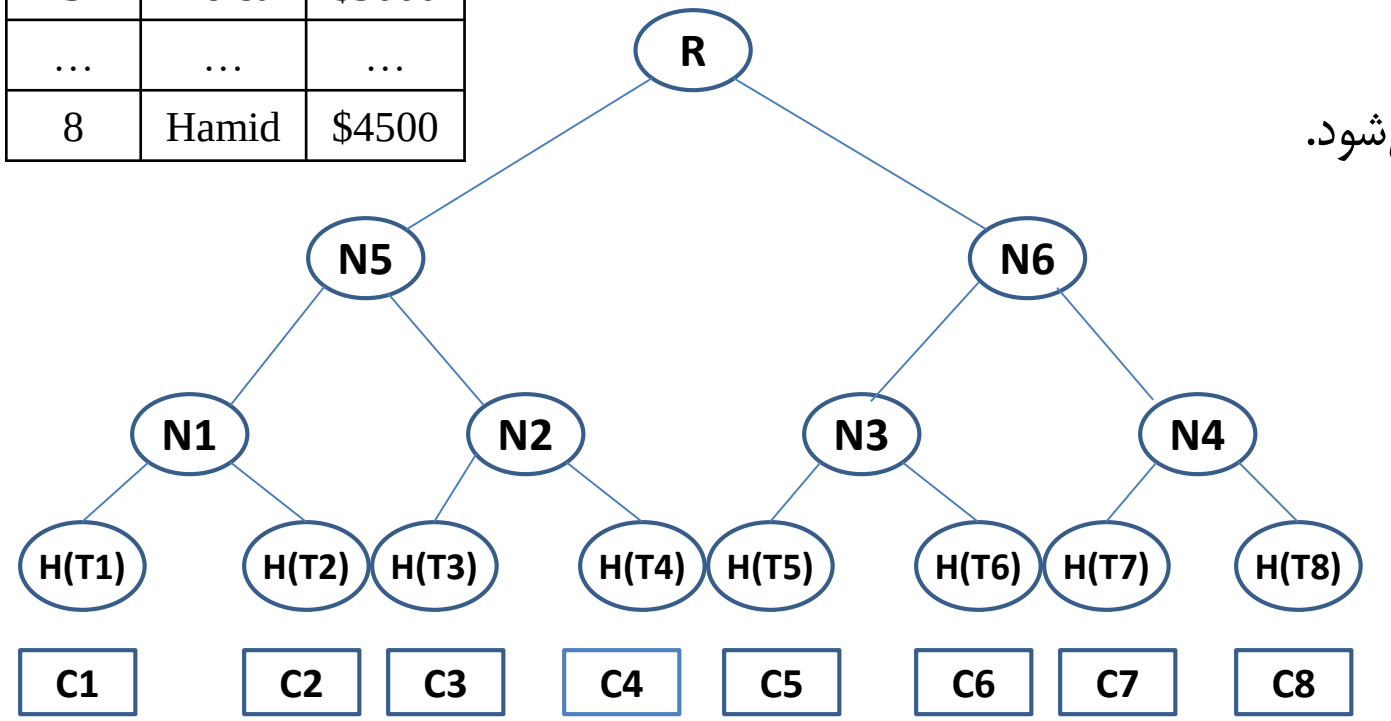
درخت چکیده ساز مرکل – معرفی



- ساخت درخت چکیده‌ساز مرکل [Devanbu 2001]
  - تاپل‌های رابطه بر اساس صفت جستجوپذیر مرتب می‌شوند.
    - ترتیب برای وارسی کامل بودن پاسخ است.
  - درخت روی مقادیر مرتب شده ایجاد می‌شود.
  - مالک داده ریشه درخت را امضا می‌کند.
  - در پاسخ به پرسمان
    - گره‌های درخت برای بازسازی ریشه علاوه بر پاسخ ارسال می‌شوند.
    - امضای ریشه توسط کارخواه وارسی می‌شود.

درخت چکیده‌ساز مرکل – مثال

Id	Name	Salary
1	Ali	$5000
2	Babak	$3500
3	Dorsa	$9000
…	…	…
8	Hamid	$4500



SELECT Name FROM Table WHERE C3 < Salary< C5

- در سناریوی برون‌سپاری
  - Ciها مقادیر صفت جستجوپذیر هستند.
  - Tiها تاپل‌های رابطه‌اند.
  - R توسط مالک داده امضا می‌شود.

$$R = H(N5 \,||\, N6)$$
$$N5 = H(N1 \,||\, N2)$$
$$N6 = H(N3 \,||\, N4)$$
$$N1 = H(H(T1) \,||\, H(T2))$$
$$N2 = H(H(T3) \,||\, H(T4))$$
$$N3 = H(H(T5) \,||\, H(T6))$$
$$N4 = H(H(T7) \,||\, H(T8))$$

- تازگی پاسخ؟

درخت چکیده‌ساز مرکل – ویژگی‌ها

- ✓ ساخت ساده شیء وارسی
- ✓ وارسی تازگی پاسخ
- ✗ شیء وارسی بزرگ
- ✗ بهنگام‌سازی داده
- ✗ وابستگی زمان وارسی به اندازه پایگاه داده

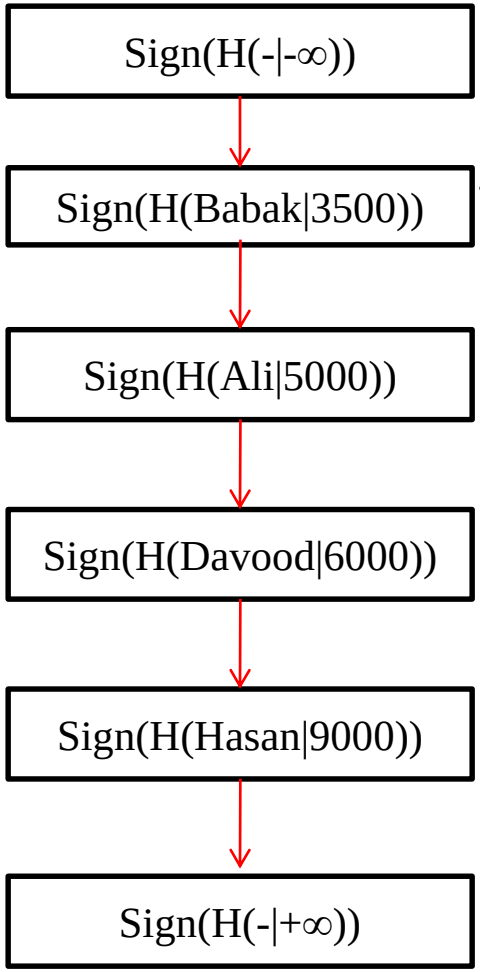
وارسی مبتنی بر امضا – معرفی

- امضا برای وارسی صحت پیام
- مالک داده در ریزدانگی مورد نظر، داده را امضا می کند.
  - ریزدانه: سربار محاسباتی
  - درشت‌دانه: سربار ارتباطی
  - امضا در سطح تاپل
- کامل بودن پاسخ؟

Name	Salary
Ali	$5000
Babak	$3500
Hasan	$9000
Davood	$6000

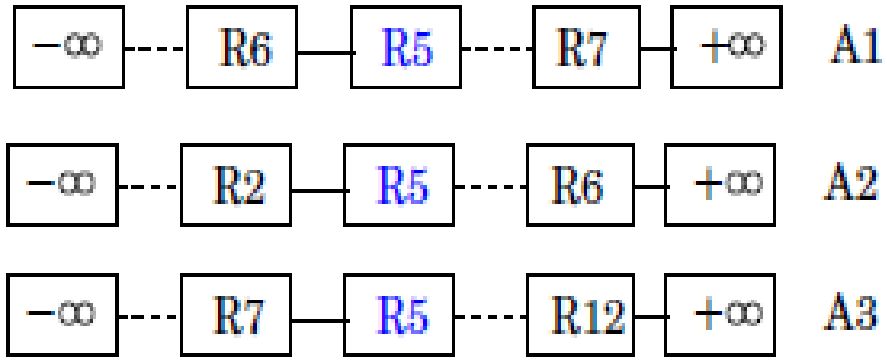
```
SELECT Name
FROM Employee
WHERE  Salary > 7000
```

(Hasan,9000)



زنجیره‌سازی امضا

- زنجیره‌سازی امضا برای وارسی کامل بودن [Narasimha and Tsudik – 2006]
  - A1، A2 و A3 بعدهای جستجوپذیر
  - تاپل‌ها بر اساس هر کدام از بعدهای جستجوپذیر مرتب می شوند.
  - $Sign(R_5) = h(h(R_5)||h(R_6)||h(R_2)||h(R_7))$



- وارسی: آیا زنجیره شکسته شده است؟
- امضای یکپارچه (aggregated signature) برای کاهش سربار ارتباطی و محاسباتی
  - Condensed RSA

$$\sigma_{aggregated} = \prod_{i=1}^{t} \sigma_i$$

ویژگی‌های روش‌های مبتنی بر امضا

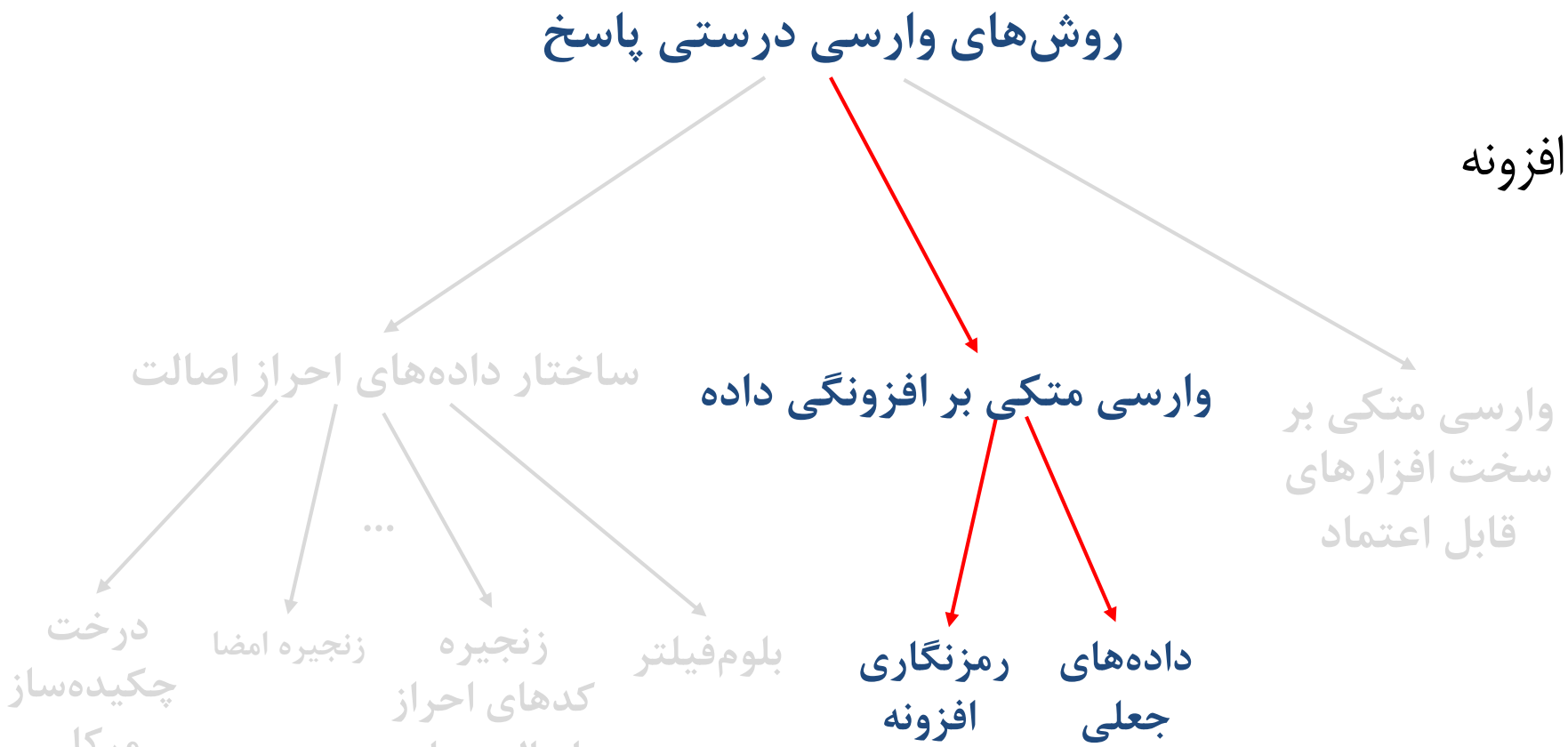
- ✓ کارایی بهنگام‌سازی
- ✓ ساختار دسترسی یکتا برای بعدهای جستجوی مختلف
- ✗ تازگی در طرح‌های مبتنی بر امضا؟

❑ مرور و مقایسه روش‌ها

- S. Bajaj, R. Sion, "**CorrectDB: SQL engine with practical query authentication**", in VLDB 2013, pp: 529-540.

روش‌های مبتنی بر افزونگی

- افزودن داده‌های جعلی
- رمزنگاری افزونه

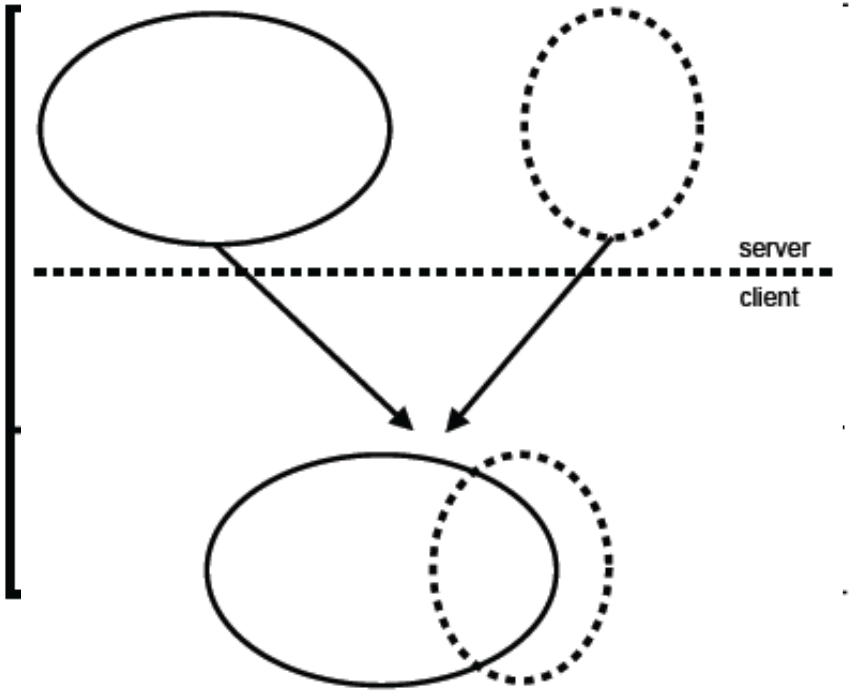


داده‌های جعلی

- افزودن تاپل‌های جعلی به پایگاه داده [Xie et al. 2007]
- تمایزناپذیری تاپل‌های جعلی نسبت به تاپل‌های واقعی
- بررسی پاسخ برگشتی
  - تعداد تاپل‌های جعلی موجود
  - تعداد تاپل‌های جعلی مورد انتظار در پاسخ
- بررسی کامل بودن پاسخ
- بده–بستان بین سربار ذخیره‌سازی و پردازش با احتمال درستی پاسخ

تکرار تاپل های رمزشده

- برخی تاپل‌ها دوبار و هر بار با کلید رمزنگاری متفاوت رمز می‌شوند [Wang et al. 2008].
- تاپل‌های رمز شده تکراری با سایر تاپل‌ها تمایزناپذیرند.
- فرایند وارسی پاسخ
- احتمال درستی پاسخ وابسته به نسبت افزونگی تاپل‌ها است
  - افزونگی بیشتر: حملات تناظر



جمع‌بندی و چالش‌های موجود

جمع بندی

- وارسی پاسخ معمولاً متکی بر اطلاعات اضافی به عنوان اطلاعات وارسی انجام می‌شود.
- فرایند وارسی باید صحت، کامل بودن و تازگی پاسخ را وارسی می‌کند.
- ساختار داده‌های احراز اصالت به‌صورت قطعی یا احتمالاتی درستی پاسخ را وارسی می‌کنند.
- فرایند وارسی معمولاً پرهزینه است.
  - پویایی داده
  - ریزدانگی وارسی

چالش‌های موجود

- پرس‌وجوهای دارای توابع تجمعی
  - SUM, AVERAGE, COUNT, ...

```
- SELECT Name FROM StudentCourse AND Course
      WHERE AVG(Grade) > 17 GROUPBY Name
```

- پرس وجوهای چندبعدی
  - زنجیره‌ی چندبعدی

```
- SELECT Name FROM StudentCourse AND Course
      WHERE  40135 < CourseID < 40138 AND StdID < 95111111
```

- یکپارچه‌سازی با روش‌های تأمین محرمانگی
- کارایی وارسی پاسخ
  - وارسی بر حسب نیاز؟
  - وارسی به‌عنوان خدمت؟

مراجع

1. P. Devanbu et al., “**Authentic Third-Party Data Publication**,” in *DBSec 2001,* pp.101–112.
2. M. Narasimha and G. Tsudik, “**Authentication of Outsourced Databases Using Signature Aggregation and Chaining**,” in *DASFAA 2006,* pp.420–436.
3. S. Bajaj, R. Sion, “**CorrectDB: SQL engine with practical query authentication**”, in VLDB 2013, pp: 529-540.
4. M. Xie et al., “**Integrity Auditing of Outsourced Data**,” in *VLDB 2007,* p.782–793.
5. H. Wang et al., “**Dual Encryption for Query Integrity Assurance**,” in *Proc. of the 17th ACM Conference on Information and Knowledge Management*, pp.863–872, ACM Press, 2008.

با تشکر از توجه شما