OFFSEC

www.offsec.ir

On

Birthday Attack

Mohammad Teimori Pabandi

teimori@ut.ac.ir

[OFFSEC www.offsec.ir WhoAmI]

Mohammad Teimori Pabandi

Computer Science Student at University of Tehran

Software Developer at Café Bazaar

Hashing 101

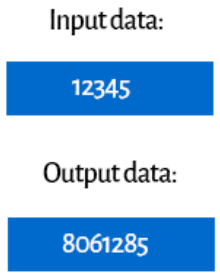
You have an important file to send. How can you make sure that is has been sent properly?

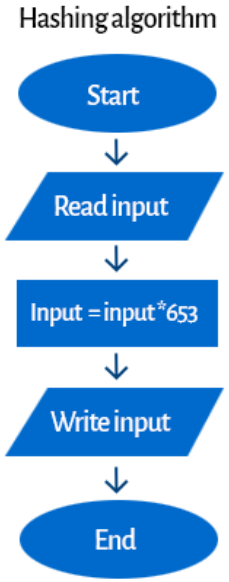
- Sending it multiple times and checking if they are identical
- Contacting the addressed and verify the file
- …

A good **hashing algorithm** can do the trick.

Hash

A **hash** is a value computed from a **base input** number using a **hashing function**.





An Example of Hash Functions

$o(x) = a \ mod \ 10$

- $o(63) = 3$
- $o(17) = 7$
- $o(27) = 7$
  - Collision!

Is this hash function secure?

The Hash Function in Your Card

The last digit in your credit card is being calculated by a hash function.

5041-7210-1208-874X

Double the green digits and hold the black ones:

[10, 0, 8, 1, 14, 2, 2, 0, 2, 2, 0, 8, 16, 7, 8]

Sum up all the digits:

(1+0) + (0) + (8) + … + (1+6) + (7) + (8) = 5**3**

The last digit will be (10 – **3**) = 7

Hashing Algorithms

A mathematical algorithm that maps data of arbitrary size to a hash of a fixed size.

Designed to be a one-way function, **infeasible to invert**.

Examples of Hashing Algorithms

MD5

- 128-bit
- MD5 (“m”) = 6f8f57715090da2632453988d9a1501b
- MD5 (“mt”) = 710998fd1b7c0235170265650770a4b1
- MD5 (“mtp”) = 5884bf0e6464833e0550ebb13e0578e3

SHA1

- 160-bit
- SHA1(“mtp”) = a2b8da9307622a2c726f73b303e6d35a22a84014

Hashing Algorithms Properties

Pre-image resistance

- Given a hash h find a message m that $h = hash(m)$

Second pre-image resistance

- Given an input m_1 find another input m_2 that $hash(m_1) = hash(m_2)$

Collision resistance

- Find two messages m_1 and m_2 that $hash(m_1) = hash(m_2)$

Birthday Attack!

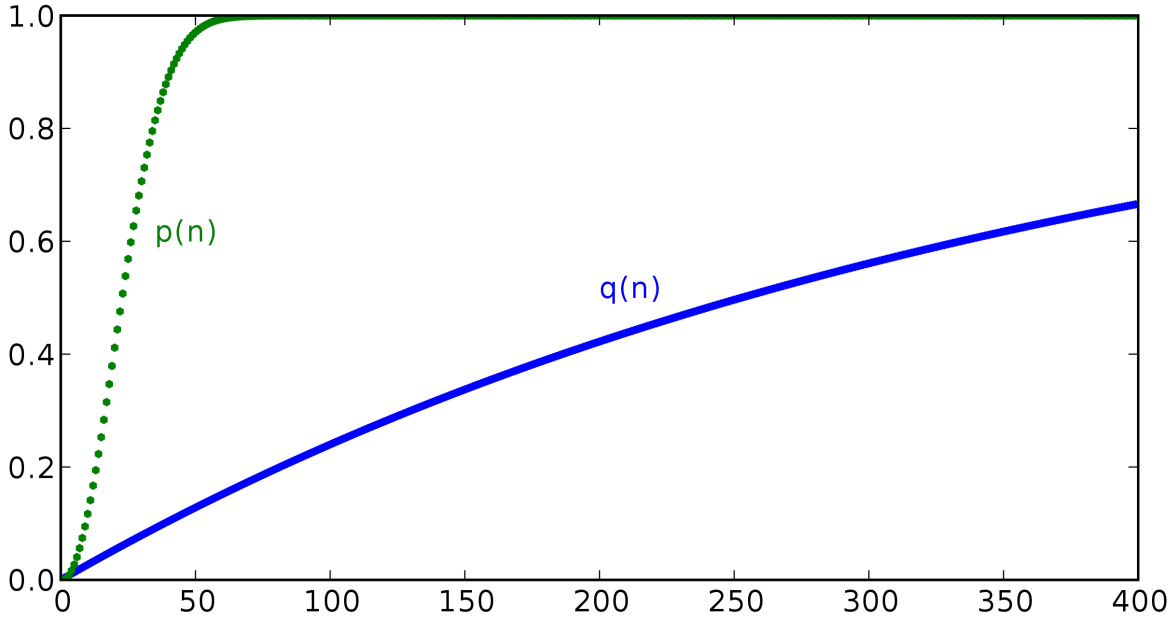
“My Birthday” Problem

My birthday is March 14. What is the probability that one specific person shares my birthday?

$$p = 1 - \frac{364}{365} = \frac{1}{365}$$

What if we chose n people?

$$q(n) = 1 - \left(\frac{365 - 1}{365}\right)^n$$



“My Birthday” Problem

$q(100) \approx 24\%$

$\boldsymbol{q(560) \approx 78.5\%}$

$q(800) \approx 89\%$

$q(1500) \approx 98\%$

$q(2000) \approx 99.5\%$

Never will be 100%!

The Birthday Problem

n random people are chosen, what is the probability that some pair of them have the same birthday?

The Birthday Problem Solved!

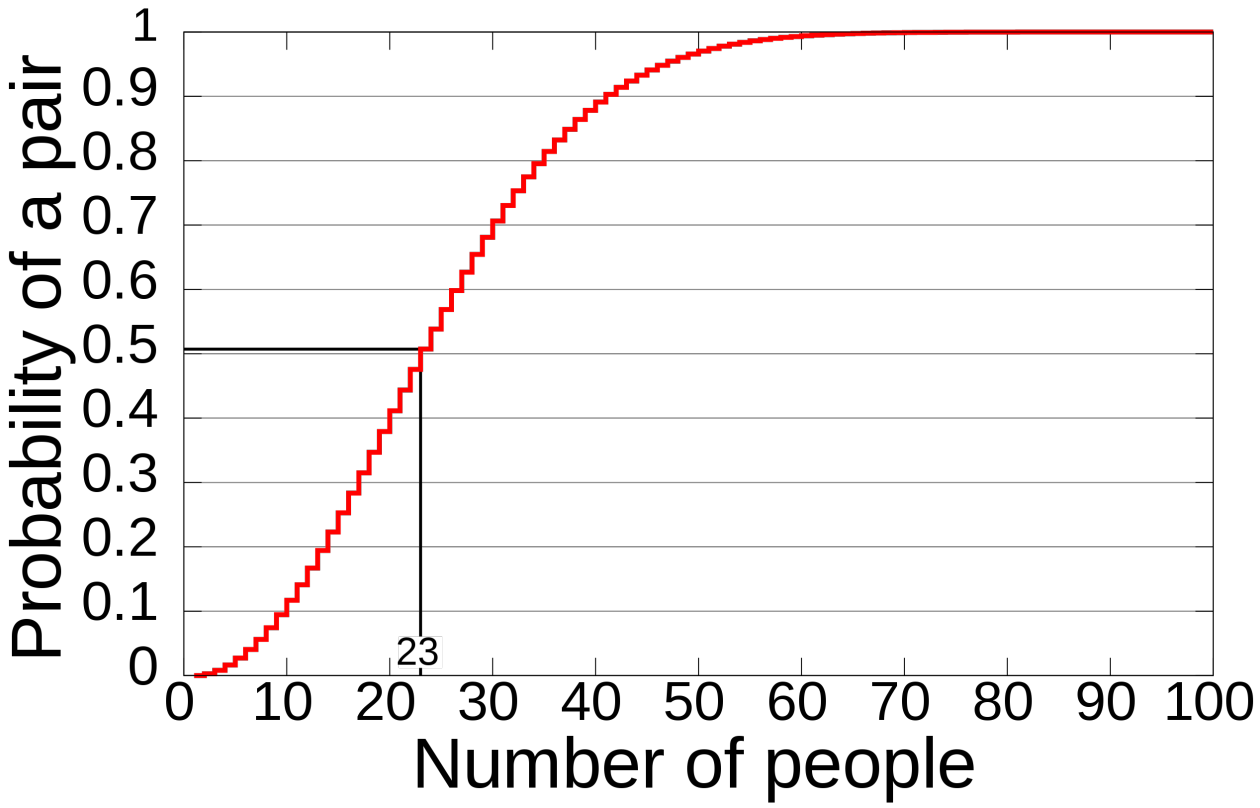
n random people are chosen, what is the probability that some pair of them have the same birthday?

$$P(n) = 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \cdots \times \left(1 - \frac{n-1}{365}\right)$$

$$= \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$$

The Birthday Problem Probability

n	$p(n)$
1	0.0%
5	2.7%
10	11.7%
20	41.1%
23	50.7%
30	70.6%
40	89.1%
50	97.0%
60	99.4%
70	99.9%
75	99.97%
100	99.999 97%
200	99.999 999 999 999 999 999 999 999 999 9998%



The Birthday Problem Probability

$P(23) \approx 50\%$

$P(70) \approx 99.9\%$

$P(366) = 100\%$

- Pigeonhole principle!

Demo time!

A Generic Attack on Collision Resistance

For $\{y_1, y_2, \dots, y_q\}$ chosen uniformly in $\{1, 2, \dots, n\}$, the probability of a collision is roughly $\frac{1}{2}$ when $q = \Theta(N^{1/2})$.

In our setting, this means that when the hash function has output length l, when taking $q = \Theta(2^{l/2})$ yields a collision with probability roughly $\frac{1}{2}$.

A Generic Attack on Collision Resistance

For a hash function to resist collision-finding attacks that run in time T, the output length of the hash function needs to be at least $2 \log T$ bits(since $2^{(2 \log T)/2} = T$).

This means that if we want finding collisions to be as difficult as an exhaustive search over 128-bit keys, then we need the output length of the hash function to be at least 256 bits.

Finding Meaningful Collisions

I wish to find two messages x and x' such that $H(x) = H(x')$.

- x is a letter from Café Bazaar explaining why I was fired.
- x' is a flattering letter of recommendation.

Then I can forge an appropriate tag on a letter of recommendation if the hash-and-MAC approach is being used by my employer to authenticate messages.

The observation is that the birthday attack only requires the hash inputs $x_1, \ldots, x_q$ to be distinct; they do not need to be random.

Finding Meaningful Collisions

I can carry out a birthday-type attack by generating $q = \Theta(2^{l/2})$ messages of the first type and q messages of the second type, and then looking for collisions between messages of the two types.

Finding Meaningful Collisions

For example:

- It is **hard/difficult/challenging/impossible** to **imagine/believe** that we will **find/locate/hire** another **employee/person** having similar **abilities/skills/character** as Mohammad. He has done a **great/super** job.
- Can be written in $4 \cdot 2 \cdot 3 \cdot 2 \cdot 3 \cdot 2 = 288$ ways!

To generate a message that can be written in 2^{64} ways, all we need is to find 64 words with one synonym each!

Or we may add punctuations.

Finding Meaningful Collisions

So I can prepare $2^{l/2}$ letters explaining why I was fired and another $2^{l/2}$ letters of recommendation; with good probability, a collision between the two types of letters will be found.

Mohammad is a **good/hardworking** and **honest/trustworthy worker/employee**.

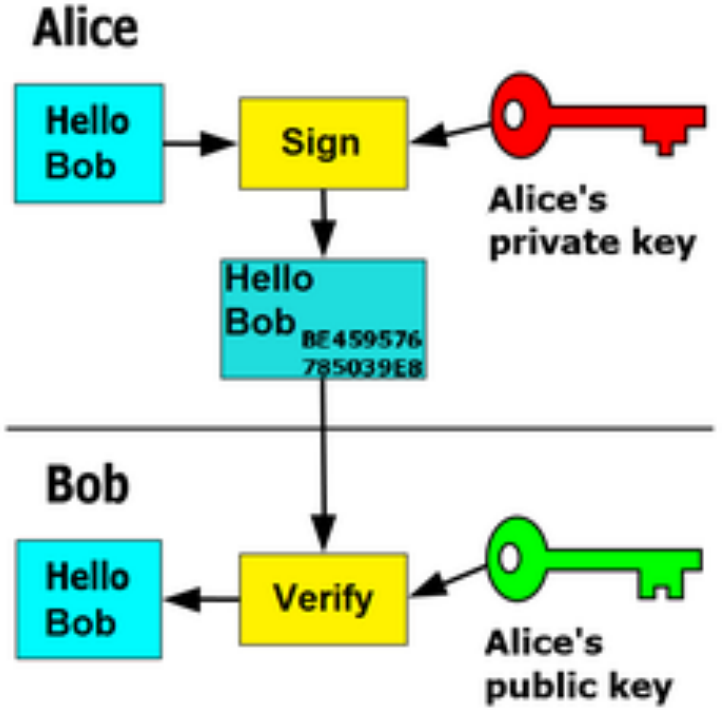
Mohammad is a **difficult/problematic** and **taxing/irritating worker/employee**.

Digital Signature

In this example the message is only signed and not encrypted.

1) Alice signs a message with her private key.

2) Bob can verify that Alice sent the message and that the message has not been modified.



Digital Signature Susceptibility

Digital signatures can be susceptible to a birthday attack. A message m is typically signed by first computing $f(m)$, where f is a cryptographic hash function, and then using some secret key to sign $f(m)$.

Mallory presents a fair version of contract to Bob for signing. After Bob has signed, Mallory takes the signature and attaches it to a fraudulent contract. This signature then "proves" that Bob signed the fraudulent contract.

Small-Space Birthday Attacks

The birthday attacks as described require a large amount of memory

- $O(q) = O(2^{l/2})$

Floyd's Cycle Finding Algorithm

Uses constant amount of memory!

ALGORITHM 5.9
A small-space birthday attack

Input: A hash function $H : \{0,1\}^* \to \{0,1\}^\ell$
Output: Distinct x, x' with $H(x) = H(x')$

$x_0 \leftarrow \{0,1\}^{\ell+1}$
$x' := x := x_0$
for $i = 1, 2, \ldots$ **do**:
  $x := H(x)$
  $x' := H(H(x'))$
  // now $x = H^{(i)}(x_0)$ and $x' = H^{(2i)}(x_0)$
  if $x = x'$ **break**
$x' := x,\ x := x_0$
for $j = 1$ to i:
  if $H(x) = H(x')$ **return** x, x' and **halt**
  else $x := H(x),\ x' := H(x')$
  // now $x = H^{(j)}(x_0)$ and $x' = H^{(i+j)}(x_0)$

Thank you for your attention!