Fundamentals of Information & Network Security
ECE 471/571

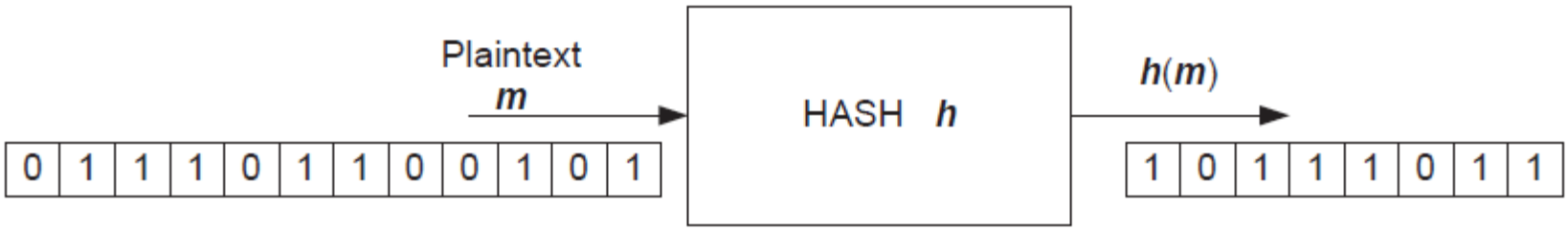
Lecture #21: Introduction to Hash Functions

Instructor: Ming Li
Dept of Electrical and Computer Engineering
University of Arizona

Cryptographic Hash Functions

- A.k.a *message digest,* one-way transformations
  - Function: mapping an arbitrary-length message to a fixed-length output (message digest)
  - One-way: impossible to reverse

- History:
  - MD2 →MD4→MD5→SHA→SHA-1 →SHA-2



Applications of Hash Functions

- Integrity Check
  - Object identifiers, File fingerprint
  - Message integrity – keyed hash (MAC)
- Source Authentication
  - Password hashing
  - Data authentication
  - Authentication Protocols
- Commitment Protocols
- Confidentiality Protection
- ……

– **How they work?**

Properties of Good Hash Functions

- Properties
  - Efficiency
    - Easy to compute *h(x)* for a given *x*.
  - One-wayness
    - Preimage resistance: Given *y*, it is computationally infeasible to compute *x* with *y=h(x)*
    - Second Preimage resistance: Given *x* and *h(x),* it is computationally infeasible to compute *x'* with *h(x)=h(x')*
  - Collision-resistance
    - It is computational infeasible to find a pair *(x, x'), $x \neq x'$* satisfying *h(x)=h(x').*

- Randomness requirement
  - For an arbitrary change in the input, every bit in the output has 50% chance to change
  - Any two outputs completely uncorrelated
  - Random oracle model

Finding the Pre-image(s)

Algorithm 4.1: FIND-PREIMAGE(h, y, Q)

choose any $\mathcal{X}_0 \subseteq \mathcal{X}, |\mathcal{X}_0| = Q$
for each $x \in \mathcal{X}_0$
  do { **if** $h(x) = y$
    then return (x)
return (failure)

Suppose that $\mathbf{Pr}[h(x) = y] = \frac{1}{M}$, for all $x \in \mathcal{X}$ and y $\in \mathcal{Y}$

- What is the average success probability of this algorithm?

$$\epsilon = 1 - \left(1 - \frac{1}{M}\right)^{Q}$$

When does this equal to ½?

Math on the Elmo…

- Finding second pre-image can be analyzed in a similar way

Finding Collision

Algorithm 4.3: FIND-COLLISION(h, Q)

choose $\mathcal{X}_0 \subseteq \mathcal{X}, |\mathcal{X}_0| = Q$
for each $x \in \mathcal{X}_0$
  do $y_x \leftarrow h(x)$
if $y_x = y_{x'}$ for some $x' \neq x$
  then return (x, x')
  else return (failure)

An Example…

Type 1 message

I am writing {this memo | } to {demand | request | inform you} that {Fred | Mr. Fred Jones} {must | } be {fired | terminated} {at once | immediately}. As the {July 11 | 11 July} {memo | memorandum} {from | issued by} {personnel | human resources} states, to meet {our | the corporate} {quarterly | third quarter} budget {targets | goals}, {we must eliminate all discretionary spending | all discretionary spending must be eliminated}.

{Despite | Ignoring} that {memo | memorandum | order}, Fred {ordered | purchased} {Post-Its | nonessential supplies} in a flagrant disregard for the company's {budgetary crisis | current financial difficulties}.

Type 2 message

I am writing {this letter | this memo | this memorandum | } to {officially | } commend Fred {Jones | } for his {courage and independent thinking | independent thinking and courage}. {He | Fred} {clearly | } understands {the need | how} to get {the | his} job {done | accomplished} {at all costs | by whatever means necessary}, and {knows | can see} when to ignore bureaucratic {nonsense | impediments}. I {am hereby recommending | hereby recommend} {him | Fred} for {promotion | immediate advancement} and {further | } recommend a {hefty | large} {salary | compensation} increase.

Birthday Paradox

- Example: if there are 23 people in a room, then what is the probability that at least two people will have the same birthday (out of 365 days in the year)?

  Larger than 0.5!

- Assume n inputs (number of people) and k possible outputs (365 days)

- If $n > k^{1/2}$, there is a good chance of finding a matching pair

- Exact math on Elmo…

- Implications to hash functions?

The Length of Hash Output

- If the digest length is *n* bits long, it takes $O(2^n)$ time to find a message with a particular pre-specified digest
- If the digest length is *n* bits long, it takes $O(2^{n/2})$ time to find two messages with the same digest (the Birthday problem)
- Because of the birthday attack, the length of hash outputs in general should double the key length of block ciphers
- SHA-256, SHA-384, SHA-512 to match the new key lengths (128,192,256) in AES

Comparison of Security Criteria

- Collision resistance implies second preimage resistance

- Collision resistance implies preimage resistance (when $|\mathcal{X}|>2^*|\mathcal{Y}|$)