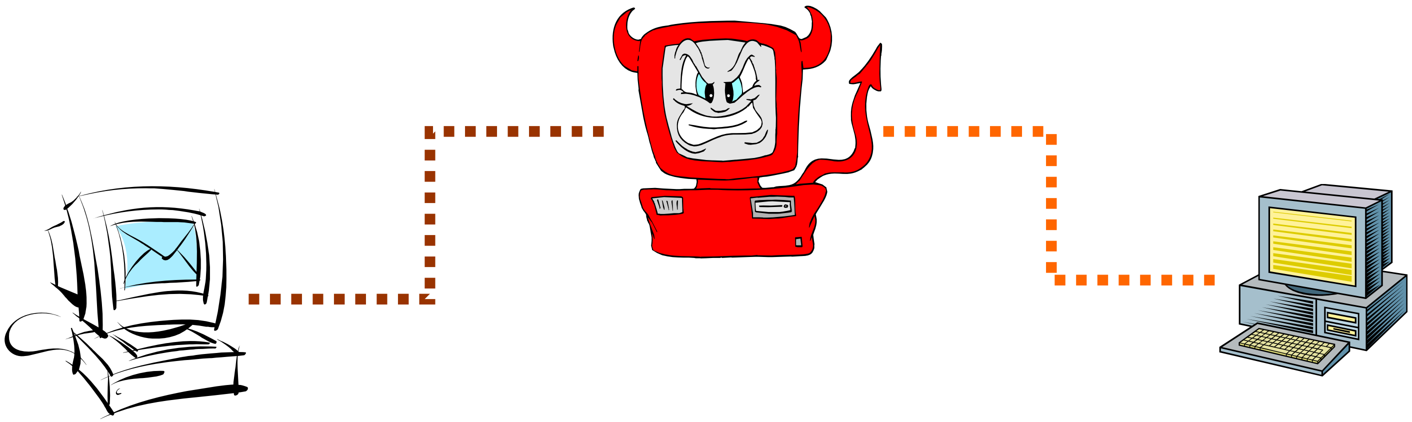
Computer Security
CS 426
Lecture 5

Cryptography: Cryptographic Hash Function

Data Integrity and Source Authentication



- Encryption does not protect data from modification by another party.
- Need a way to ensure that data arrives at destination in its original form as sent by the sender and it is coming from an authenticated source.

Cryptographic Hash Functions

- A hash function maps a message of an arbitrary length to a m-bit output
  - output known as the fingerprint or the message digest
  - if the message digest is transmitted **securely**, then changes to the message can be detected

- A hash function is a many-to-one function, so collisions can happen.

Security Requirements for Cryptographic Hash Functions

Given a function $h:X \rightarrow Y$, then we say that h is:

- preimage resistant (one-way):

  if given $y \in Y$ it is computationally infeasible to find a value $x \in X$ s.t. $h(x) = y$
- 2-nd preimage resistant (weak collision resistant):

  if given $x \in X$ it is computationally infeasible to find a value $x' \in X$, s.t. $x' \neq x$ and $h(x') = h(x)$
- collision resistant (strong collision resistant):

  if it is computationally infeasible to find two distinct values $x', x \in X$, s.t. $h(x') = h(x)$

Uses of hash functions

- Software integrity
  - E.g., tripwire
- Timestamping
  - How?

- Message authentication
- One-time Passwords
- Digital signature

Bruteforce Attacks on Hash Functions

- Attacking one-wayness
  - Goal: given $h: X \to Y$, $y \in Y$, find x such that $h(x)=y$
  - Algorithm:
    - pick a random value x in X, check if h(x)=y, if h(x)=y, returns x; otherwise iterate
    - after failing q iterations, return fail
  - The average-case success probability is

$$\varepsilon = 1 - \left(1 - \frac{1}{|Y|}\right)^q \approx \frac{q}{|Y|}$$

  - Let $|Y|=2^m$, to get ε to be close to 0.5, $q \approx 2^{m-1}$

Bruteforce Attacks on Hash Functions

- Attacking collision resistance
  - Goal: given h, find x, x' such that h(x)=h(x')
  - Algorithm: pick a random set X_0 of q values in X
    for each $x \in X_0$, computes $y_x = h(x)$
    if $y_x = y_{x'}$ for some $x' \neq x$ then return (x,x') else fail
  - The average success probability is $1 - e^{-\frac{q(q-1)}{2|Y|}}$
  - Let $|Y| = 2^m$, to get ε to be close to 0.5, $q \approx 2^{m/2}$
  - This is known as the birthday attack.

Well Known Hash Functions

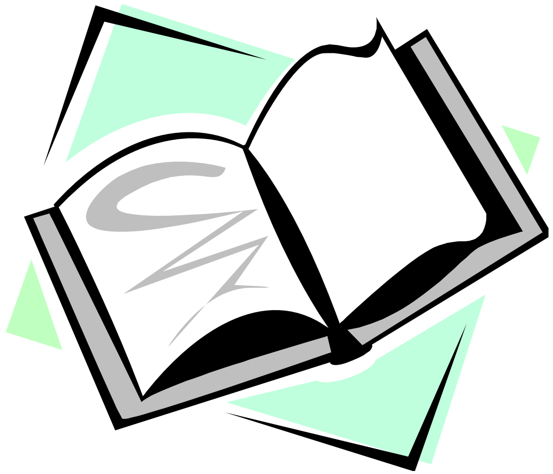
- MD5
  - output 128 bits
  - collision resistance completely broken by researchers in China
- SHA1
  - output 160 bits
  - no collision found yet, but method exist to find collisions in less than 2^80
  - considered insecure for collision resistance
  - one-wayness still holds
- SHA2 (SHA-224, SHA-256, SHA-384, SHA-512)
  - outputs 224, 256, 384, and 512 bits, respectively
- NIST is having an ongoing competition of new standard hash algorithms, 14 algorithms currently considered

Choosing the length of Hash outputs

- Because of the birthday attack, the length of hash outputs in general should double the key length of block ciphers
  - SHA-224 matches the 112-bit strength of triple-DES
  - SHA-256, SHA-384, SHA-512 match the new key lengths (128,192,256) in AES

Readings for This Lecture

- Wikipedia
  - Cryptographic Hash Function



Coming Attractions …

- Cryptography: Message Authentication Code.