Foundations of Computer Security

Lecture 50: Cryptographic Hash Functions

Dr. Bill Young
Department of Computer Sciences
University of Texas at Austin

Hash Functions

A *hash function* is a function that converts variable-sized text into a small datum, usually a fixed size integer.

A *cryptographic hash function* has the additional qualities:

- it is difficult to construct a text that has a given hash,
- it is difficult to modify a given text without changing its hash,
- it is unlikely that two different messages will have the same hash.

The hash value is sometimes called a *message digest*.
Cryptographic hash functions are used to protect integrity.

Vocabulary

A function f is *preimage resistant* if, given h, it is hard to find any m such that $h = f(m)$.

A function f is *second preimage resistant* if, given an input m_1, it is hard to find $m_2 \neq m_1$ such that $f(m_1) = f(m_2)$. This is sometimes called *weak collision resistance*.

A function f is (strong) *collision resistant* if it is hard to find two messages m_1 and m_2 such that $f(m_1) = f(m_2)$.

Birthday Attacks

If a function $f(x)$ yields any of H different outputs with equal probability and H is sufficiently large, then we expect to obtain a pair of different arguments x_1 and x_2 with $f(x_1) = f(x_2)$ after evaluating the function for about $1.25\sqrt{H}$ different arguments on average.

What does this mean for a hash value of 128 bits? for 160 bits?

Cryptographic Hash Functions

Hash functions usually are used for integrity, not confidentiality.

- In a document retrieval system containing legal records, it may be important to know that the copy retrieved is identical to that stored.
- In a secure communications system, the correct transmission of messages may override confidentiality concerns.

A cryptographic hash function "binds" the bytes of a file together in a way that makes any alterations to the file apparent. We say that we *seal* the file to make it tamper-proof (actually tamper-resistant).

Using a Hash Functions

The process is as follows:

- Given a sensitive file f, compute the hash function $h(f)$ and store the result securely.
- Each time the file is used or accessed, recompute the hash.
- Compare it to the stored value.

If the two values match, it is likely that no changes have occurred to the file.

Common Hash Algorithms

Two widely used cryptographic hash functions are:

MD5: (Message Digest 5) invented by Ron Rivest and RSA Labs;

SHA-1/SHA-2/SHS: (Secure Hash Algorithm or Standard) similar to MD5.

MD5 hashes a message of any size to a 128-bit digest. SHA/SHS produce a 160-bit digest.

Lessons

- A cryptographic hash function takes an arbitrary text and produces a fixed size bit string that depends on each value of the text.
- It should be difficult to find collisions—values that hash to the same result.
- A hash can be used to show with high probability that a text has not changed.

Next lecture: Key Exchange