Introductory Cryptography

Message Digests

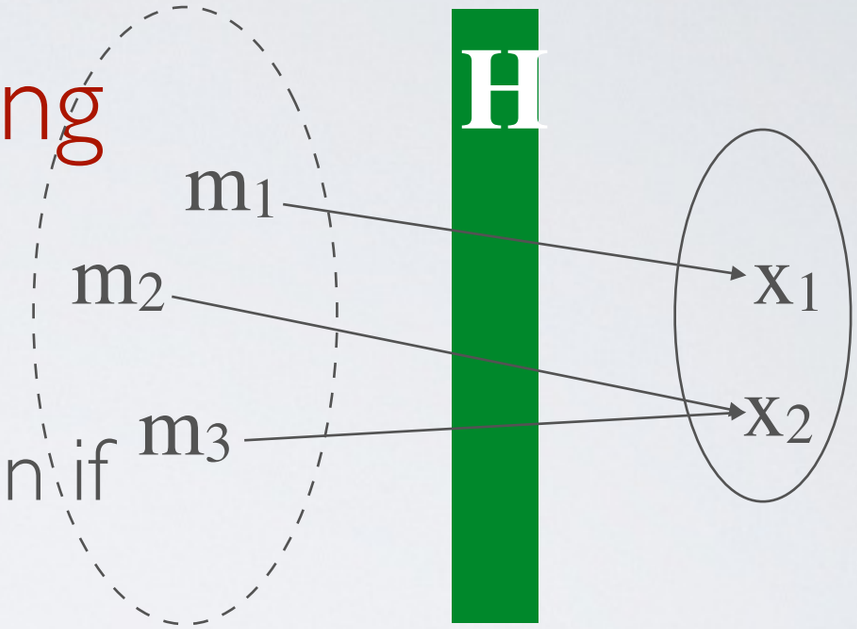
Kc Udonsi

Message digests

Message digests are meant for creating fingerprints of messages

- Un-keyed message digest : hashes, checksum
- Keyed message digests : MACs

Cryptographic hashing



$H(m) = x$ is a hash function if

- H is one-way function
- m is a message of any length
- x is a message digest of a fixed length

➡ H is a lossy compression function
necessarily there exists x, m_1 and m_2 | $H(m_1) = H(m_2) = x$

Computational complexity

- Given $\mathbf{H}$ and $\mathbf{m}$, computing $\mathbf{x}$ is **easy** (polynomial or linear)

- Given $\mathbf{H}$ and $\mathbf{x}$, computing $\mathbf{m}$ is **hard** (exponential)

➡ $\mathbf{H}$ is **not invertible**

Preimage resistance and collision resistance

PR - Preimage Resistance (a.k.a One Way)

➡ given H and x, hard to find m
e.g. password storage

2PR - Second Preimage Resistance (a.k.a Weak Collision Resistance)

➡ given H, m and x, hard to find m' such that $H(m) = H(m') = x$
e.g. virus identification

CR - Collision Resistance (a.k.a Strong Collision Resistance)

➡ given H, hard to find m and m' such that $H(m) = H(m') = x$
e.g. digital signatures

CR → 2PR and CR → PR

Security of hash functions

Brute-forcing a hash function

CR - Collision Resistance

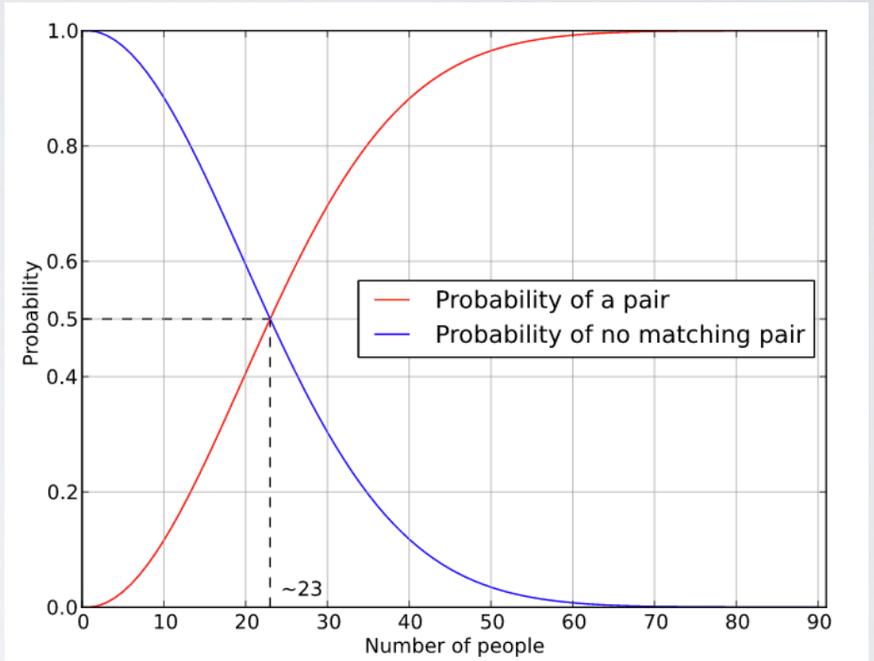
➡ given $\mathbf{H}$, hard to find $\mathbf{m}$ and $\mathbf{m'}$ such that $\mathbf{H(m) = H(m')}$ $\mathbf{= x}$

Given a hash function $\mathbf{H}$ of $\mathbf{n}$ bits output

- Reaching all possibilities — 2^n cases
- On average, an attacker should try half of them — 2^{n-1} cases

Birthday Paradox

"There are 50% chance that 2 people have the same birthday in a room of 23 people"



N-bits security

- Given a hash function **H** of **n** bits output, a collision can be found in around $2^{n/2}$ evaluations
  e.g SHA-256 is 128 bits security

Broken hash functions beyond the birthday paradox

	Year	Collision
MD5	2013	2^{24} evaluations (2^{39} with prefix)
SHA-1	2015	2^{57} evaluations