Cryptographic hash functions

Jed Crandall, jedimaestro@asu.edu

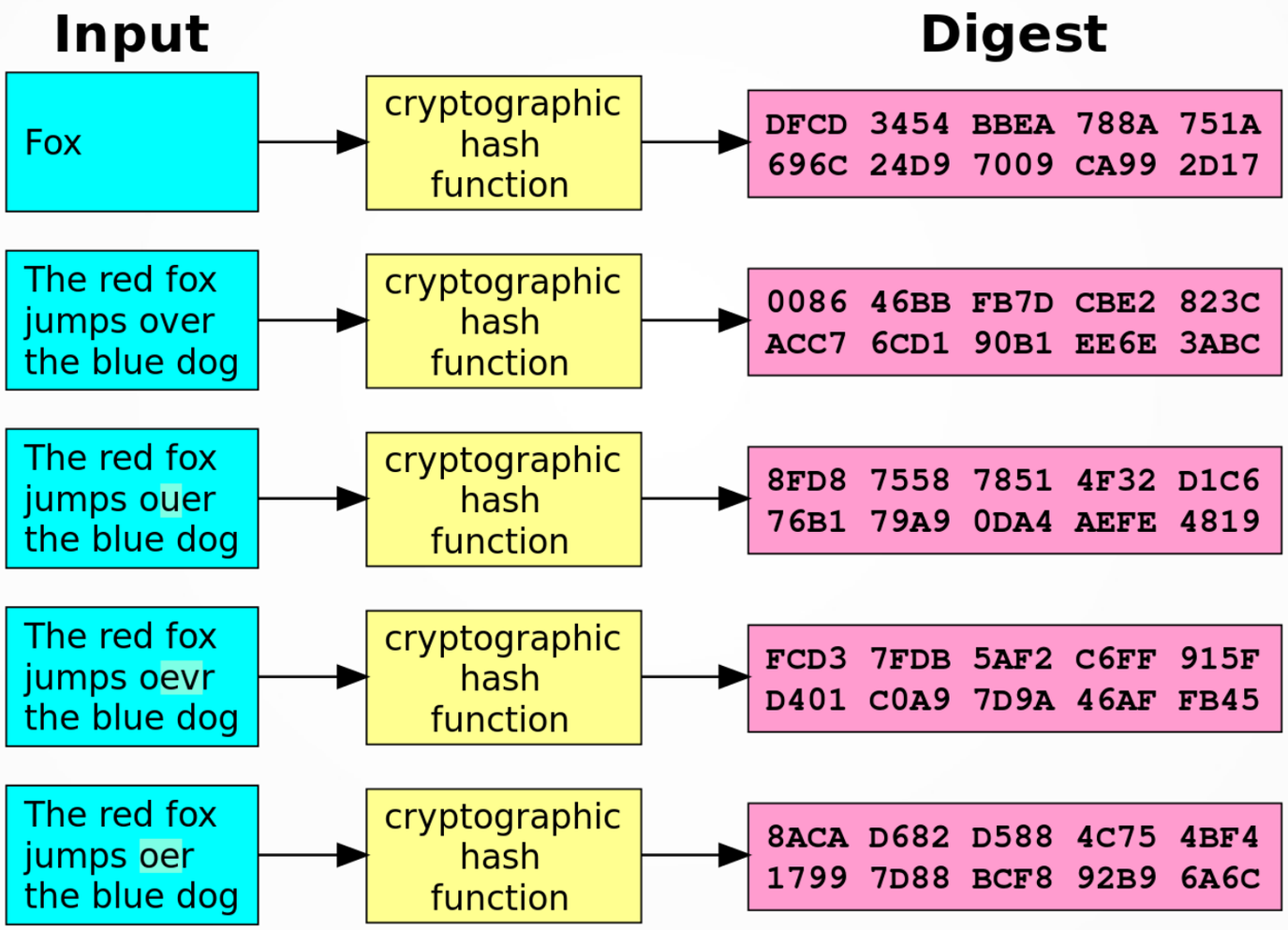
Why hash functions?

- Speed
- Error detection (*e.g.*, checksum)
- Security and privacy

Why cryptographic hash functions?

- Unique identifier for an object
- Integrity of an object
- Digital signatures
- Passwords
- Proof of work

Example



What makes a hash function cryptographic?

- One-way function
- Deterministic (same input, same output)
- Infeasible to find message that digests to specific hash value
- Infeasible to find two messages that digest to the same hash
- Avalanche effect (small change in message leads to big changes in digest---digests seemingly uncorrelated)
- Quick

Algorithms

- MD5: 128-bit digest, seriously broken
- SHA-1: 160-bit digest, not secure against well-funded adversaries
- SHA-3: 224 to 512 bit digest, adopted in August of 2015
- CRC32: not cryptographic, very poor choice

Example from the Citizen Lab report….

Property #1

- Pre-image resistance
- Given *h*, it should be infeasible to find *m* such that *h = hash(m)*

Property #2

- Second pre-image resistance
- Given a message m_1, it should be infeasible to find another message m_2 such that...
  $hash(m_1) = hash(m_2)$

Property #3

- Collision resistance
- It should be infeasible to find two messages, m_1 *and* m_2 such that...
  $hash(m_1) = hash(m_2)$

Attacks

- Pre-image attack
- Collision attack
- Chosen-prefix collision attack
- Birthday attack

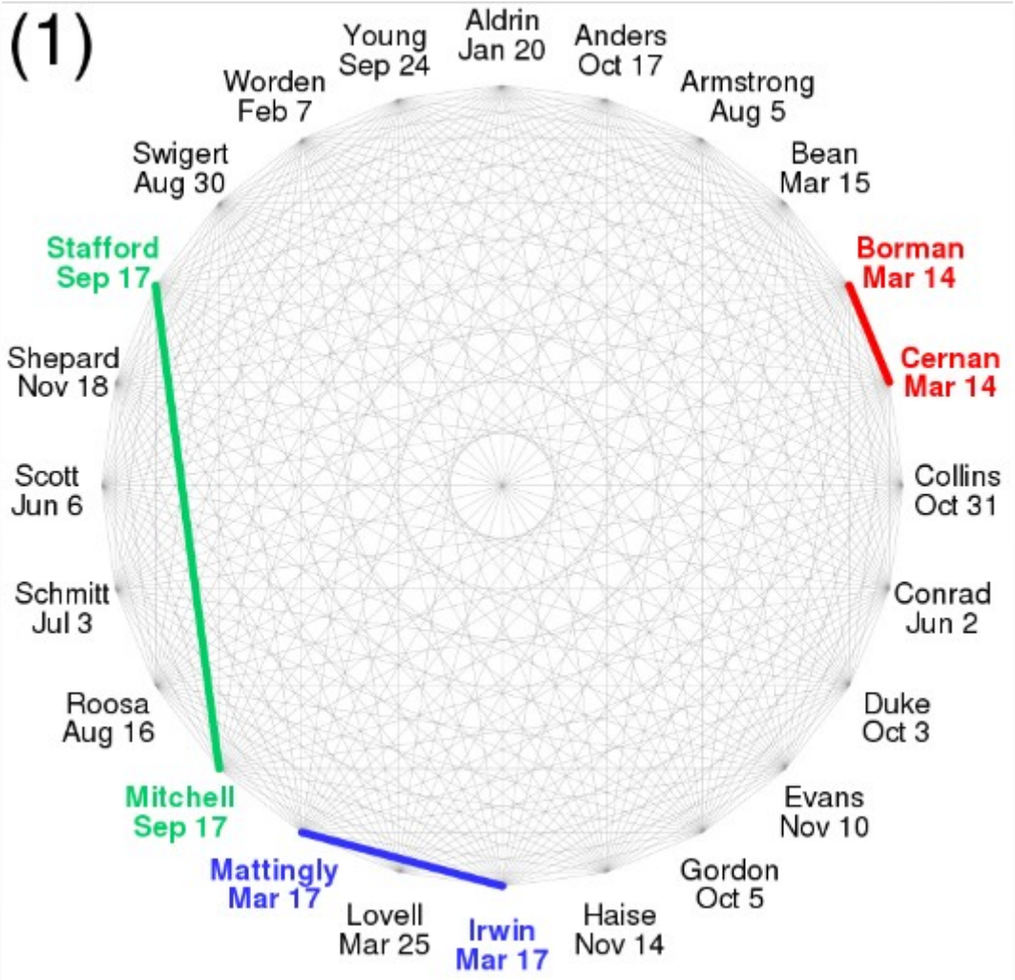
Chosen-prefix collision attack

- Given two prefixes p_1 and p_2, find m_1 and m_2 such that $hash(p_1||m_1)=hash(p_2||m_2)$
- p1 and p2 could be domain names in a certificate, images, PDFs, *etc. …* any digital image.
- This is one of the two ways MD5 is broken (other is plain old collision resistance), and is how we generated the two images with the same MD5 sum for the example from the Citizen Lab report

Birthday attack

- Probability of collision is *1* in 2^n, but the expected number of hashes until two of them collide is $sqrt(2^n)=2^{n/2}$
  - Why? Third try has two opportunities to collide, fourth has three opportunities, fifth has six, and so on...

24 people, same birthday?



https://commons.wikimedia.org/wiki/File:Birthday_attack_vs_paradox.svg

Think of a "random" 4-digit number

- *lg(9999)* is about 14, so a 14-bit number
- $sqrt(2^{14})=2^{7}=128$
- You're going to say it out loud
  - We'll go around the room, go fast
  - Don't use your bank PIN, *etc.*
  - Raise your hand and yell if someone says your number