Hash Functions

Review of desirable properties
Applications (contd.)
Construction (MD5)

Desirable Properties

1. OW "one-way" (pre-image resistance)
2. CR Collision resistance (strong)
3. TCR Target collision resistance (weak)
4. PRF Pseudo-randomness
5. NM Non-malleability

Given $h(x)$, should not be able to find $h(x+t)$. NM $\Rightarrow$ OW
Applications (contd.)

4. Commitments

Alice has value x (e.g., auction bid)
Alice then computes \( CC(x) \) and submits it as her bid

\( CC(x) \) is her "sealed bid"

When bidding is over, Alice "opens" \( CC(x) \) to reveal x

Binding: Alice should not be able to open \( CC(x) \) in multiple ways.
Secrecy: Auctioneer seeing \( CC(x) \) should not learn anything about x
NM: Given \( CC(x) \) shouldn't be possible to produce \( CC(x-1) \)

Need: NM, CR, OW (really need more for secrecy!) \( h(x) = h(x) \parallel \text{msb}(x) \)

How: \( CC(x) = h(r \parallel x) \) \( r \in \mathbb{R} \{0,1\}^{256} \)
to open reveal r & x

randomized
Merkle tree

Authenticate n objects (e.g., time-stamping)

```
X_1 X_2 X_3 X_4  # data blocks
```

```
x = h(y || z)
```

root is authenticator for all n values (put in New York Times)

Show leaf & ancestors & their siblings to prove leaf is in tree.

Need: CR
Construction ("Merkle-Damgard" style)

- Choose output size \( d \) (e.g., \( d = 160 \))
- Choose chaining variable \( c \) (e.g., \( c = 160 \))
  better if \( c \geq 2d \)
- Choose block size \( b \) for message
- Design "compression fn" \( f \)
  \[ f : \{0, 1\}^c \times \{0, 1\}^b \rightarrow \{0, 1\}^c \]
- Choose \( c \)-bit (initialization vector)
- Pad message so \( m \)'s new length is multiple of \( b \) bits

Padding:
- hash input 0000 pad with 0000
- hash input 00 00 pad with 00

Collision on two (or more) inputs.
Solution: Include length of original \( m \) in pad
Observations

IV is arbitrary, but fixed

Thm: If $f$ is CR, so is $h$.

Pf: Work backwards through chain from $h$-collision to find $f$-collision.

Thm: Same for OW.

\[ M_i \xrightarrow{f} C_i \quad \Rightarrow \quad M_i \xrightarrow{En_{C_i}} C_i \]

AES etc, hard to change keysize
TYPICAL COMPRESSION FUNCTION (MD5)

- chaining variable & output are 128 = 4 x 32 bits
- IV = fixed value
- 64 rounds; each modifies state (in reversible way) based on selected message word
- message block \( b = 512 \) bits considered as 16 32-bit words
- uses end-around XOR

\[
g(x, y, z) = \begin{cases} 
  xy \\ xz \\ y \oplus x \oplus z \\
\end{cases}
\]

depending on round