Outline

First talk:

- Background on hash function
- Previous work on SHA-0, and SHA-1
- Improved collision search attack on SHA-0
    - Brief description

Second talk:

- Collision search attack on SHA-1
    - Major steps, with focus on intuition
- Summary

Efficient Collision Search Attacks on SHA-0

CRYPTO 2005

August 15, 2005

Xiaoyun Wang

Hongbo Yu

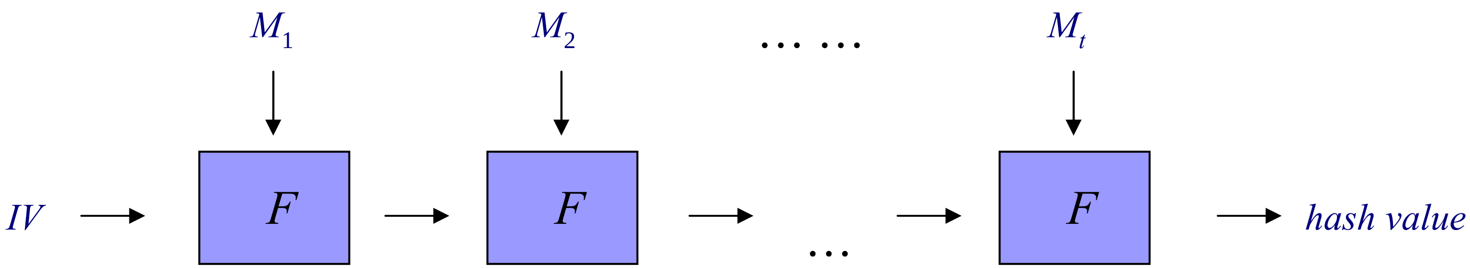
Yiqun Lisa Yin

Hash functions

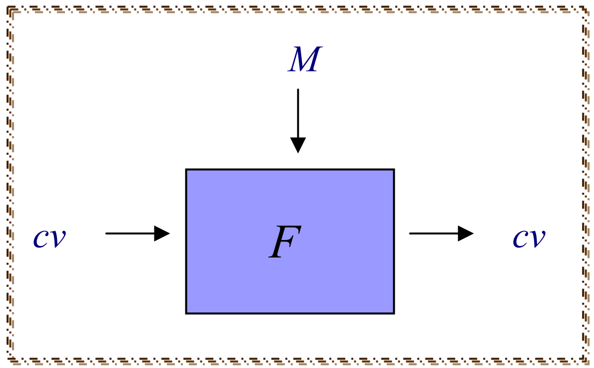
- *Cryptographic* hash function: $y = h(x)$
  - Take a message x of arbitrary length
  - Output a short value y of a fixed length
    - y is called *hash value* or *message digest*
- Basic security properties
  - One-way: given y, hard to find x s.t. $x = h^{-1}(y)$
  - Collision resistant: hard to find $x \neq y$ s.t. $h(x) = h(y)$
- Applications
  - Digital signatures, password verification, key generation …
  - Present in almost all security systems

General design approach

- Iterative structure
  - Input message is divided into *fixed-length* blocks
  - Each block is processed using a *compression function F*



- Design of the compression function
  - Block-cipher based
  - Customized design "from scratch"
    - the MDx family

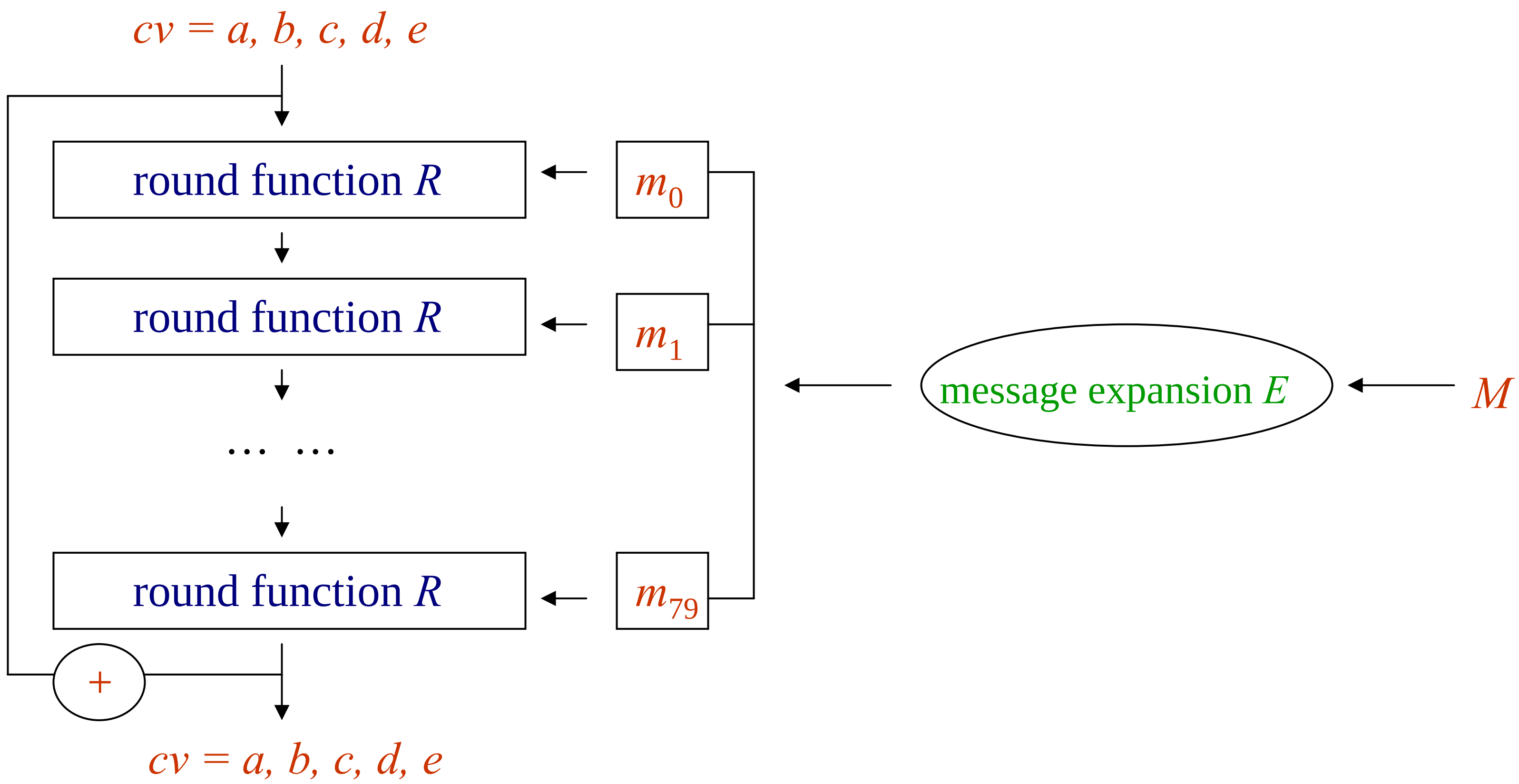


The MDx family of hash functions

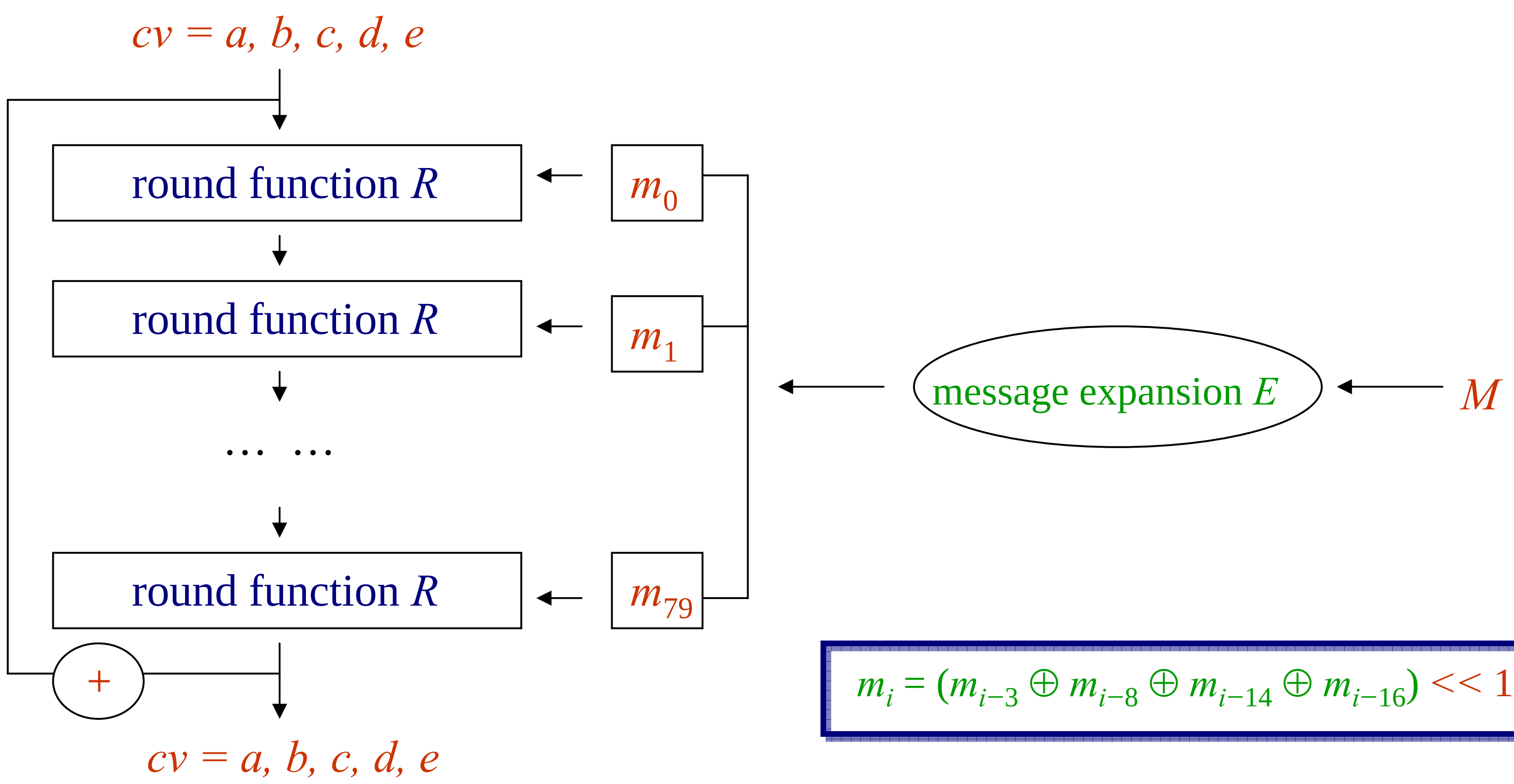
- Design philosophy
  - Using simple operations available on modern computers
    - Easy implementation, good performance
- Most popular ones
  - MD4 (1990)
  - MD5 (1991)

  - SHA-0 (1993)
  - SHA-1 (1995)
  - SHA-2 (2001)
    - SHA-224, 256, 384, 512

Compression function of SHA-0 & SHA-1



Compression function of SHA-0 & SHA-1



$$m_i = (m_{i-3} \oplus m_{i-8} \oplus m_{i-14} \oplus m_{i-16}) << 1$$

SHA-0 doesn't have <<1

R:

$$a = (a << 5) + f(b, c, d) + e + m_i + k_i$$

$$a \rightarrow b \rightarrow (<< 30) \rightarrow c \rightarrow d \rightarrow e$$

Security strengths

- Expected security level
  - Depends on hash output length n
  - One-way: 2^n
  - Collision resistant: $2^{n/2}$
- Security of MDx against collision search attacks

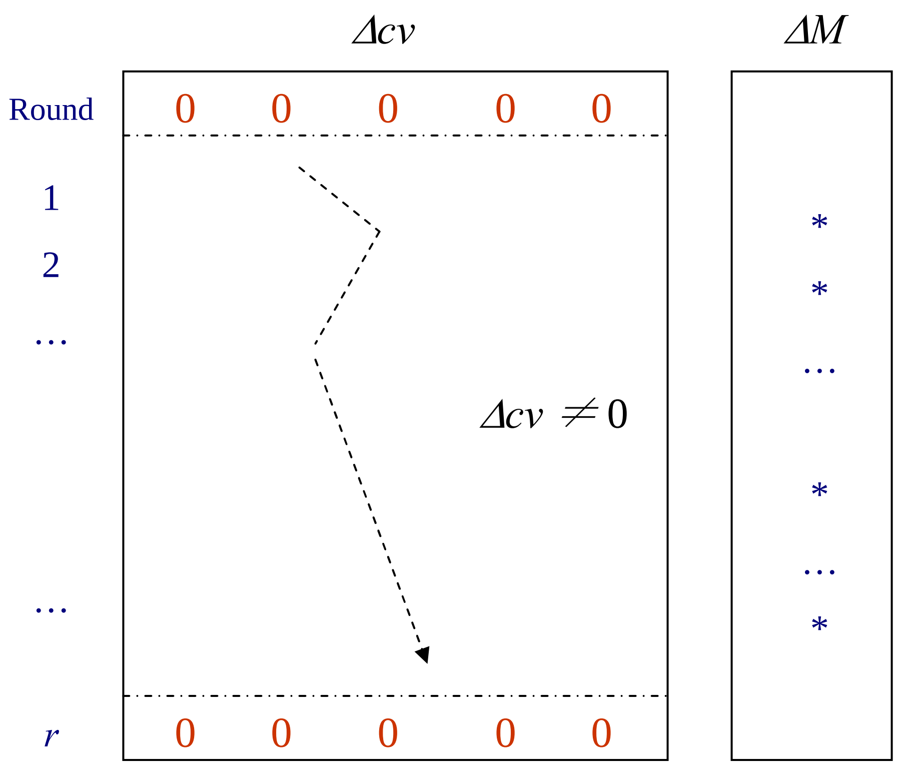
Hash function	Expected strength	Best known collision attack
MD4	2^{64}	~ 3
MD5	2^{64}	~ 2^{30+}
SHA-0	2^{80}	**2^{39}**
SHA-1	2^{80}	**2^{69}**
SHA-256	2^{128}	?

Our new results

Previous work on SHA-0 & SHA-1

- Chabaud and Joux (Crypto'98)
  - Collision attack on SHA-0, with complexity 2^{61}
  - Introduced two useful tools: local collision and disturbance vector
- Wang (Manuscripts, 97 – 98)
  - Independent analysis on SHA-0
  - Message modification techniques and algebraic analysis
- Biham and Chen (Crypto'04)
  - Near collision attack on SHA-0, with complexity 2^{40}
  - Neutral bit techniques
- Biham, Chen, Joux etc. (Crypto'04 Rump, Eurocrypt'05)
  - First real collision of SHA-0 found, with complexity 2^{51}
  - Collision attack on SHA-1 reduced to 50+ rounds
  - Multi-block techniques
- Rijmen and Osward (RSA-CT'05)
  - Collision attack on SHA-1 reduced to 53 rounds
  - Analysis using insight from coding theory

Overview of a collision attack: differential style attacks



- Differential attacks were first introduced to analyze block ciphers
- Basic ideas applicable to hash functions
  - difference: $\Delta x = x \oplus x'$
  - Express a collision of *F*

    $$\Delta cv = 0,\ \Delta M \neq 0 \ \rightarrow\ \Delta cv = 0$$

  - Differential path
    - Intermediate differences
    - Holds with some prob *p*
    - Complexity is about $1/p$

Chabaud and Joux's Attack on SHA-0

- Basic idea
  - Find *local collision* — a collision spanning a few steps
    - By analyzing round function
  - Stack local collisions together to form a global collision
    - By analyzing message expansion

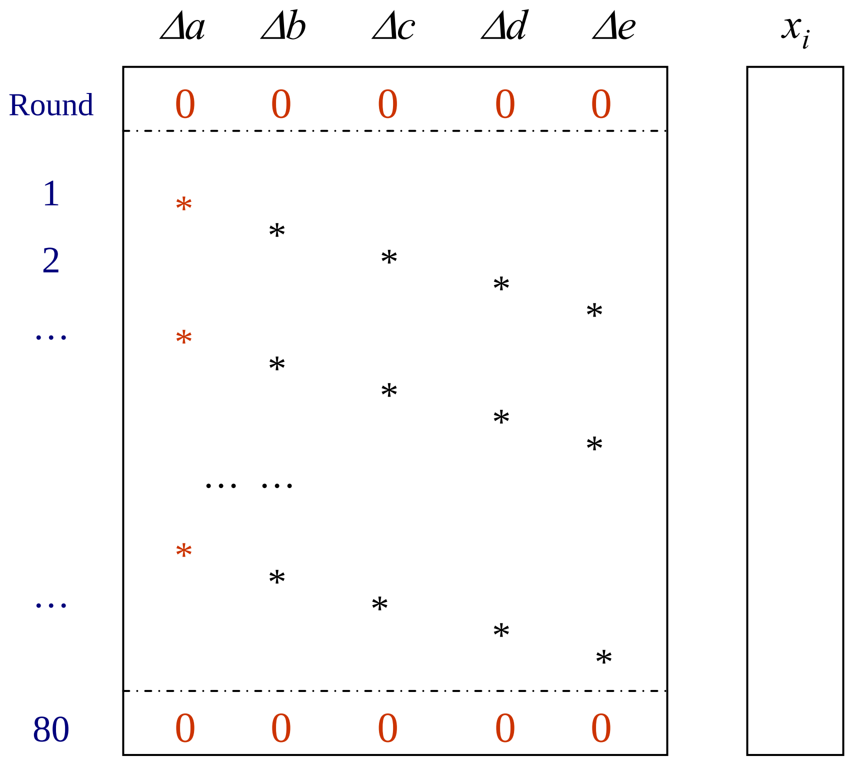
Local collision of SHA-0

Local collision: a 6-round diff path with $\Delta cv = 0$ before and after.

round	Δm_{i-1}	Δa_i	Δb_i	Δc_i	Δd_i	Δe_i
i-1		0	0	0	0	0
i	*	*				
i+1	*		*			
i+2	*			*		
i+3	*				*	
i+4	*					*
i+5	*	0	0	0	0	0

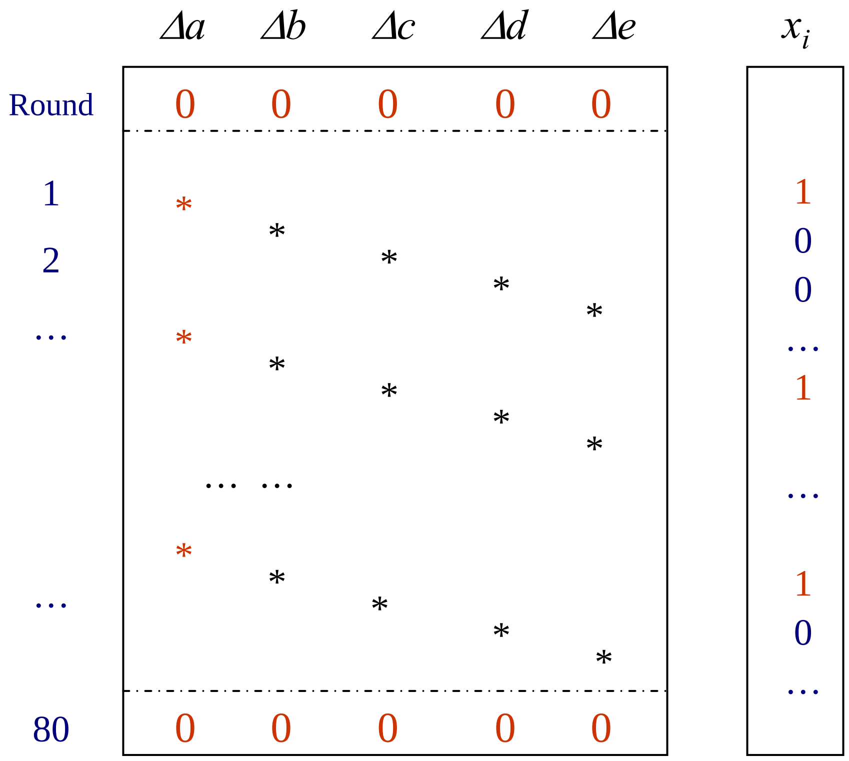
Local collision can start at any round. *Probability* is about $2^{-2} – 2^{-5}$.

Disturbance vector (DV) for SHA-0



- Stack local collisions
  - Need to specify starting points of local collisions

Disturbance vector (DV) for SHA-0



- DV: $x_0, x_1, x_2, \ldots, x_{79}$
  - $x_i = 1$ iff a local collision starts in round i
- DV satisfies message expansion
  - Possible choices: 2^{16}
- Three conditions on DV
  - Ensure that local collisions can be put together properly
  - Only 3 vectors left
- What is a good DV?
  - low Hamming weight

  → Higher prob for the path

Improved collision search attack on SHA-0 — brief description

1. Construct differential path
   - Select a good DV
     - Search in less constraint vector space (fewer conditions)
     - HW of the DV is lower than those in existing attacks
   - Fine tune the differential path
2. Boost success probability of the attack
   - Apply message modification techniques from the attack on MD5

- Complexity of the attack: 2^{39}
  - Real collisions can be found quickly

Finding Collisions in the Full SHA-1

CRYPTO 2005

August 15, 2005

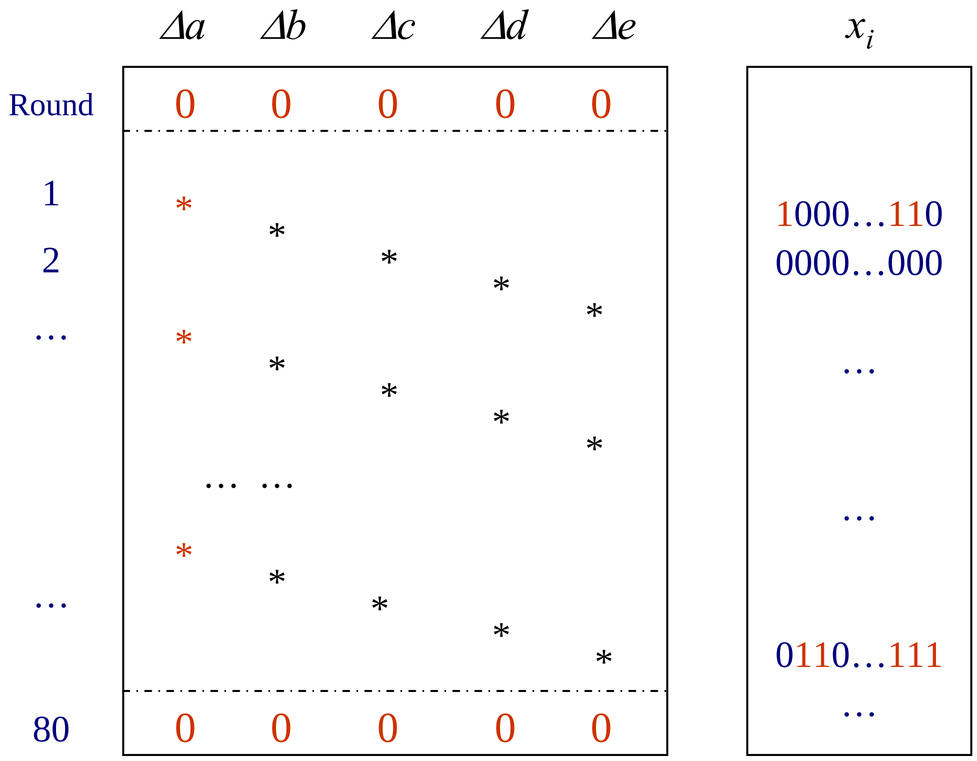
Xiaoyun Wang
Yiqun Lisa Yin
Hongbo Yu

Collision search attack on SHA-1

1. Construct differential path
   - Leverage on techniques from the attack on SHA-0
     - Local collisions and generalized disturbance vectors
   - Search for low Hamming weight vectors
     - Exploit weakness in SHA-1 message expansion
   - Fine tune the differential path
     - Exploit weakness in round function

2. Boost success probability of the attack
   - Apply techniques from the attack on MD5
     - Derive conditions associated with the differential path
     - Modify messages so that many of the conditions hold with *probability one*
     - Construct two-block collision using near collision

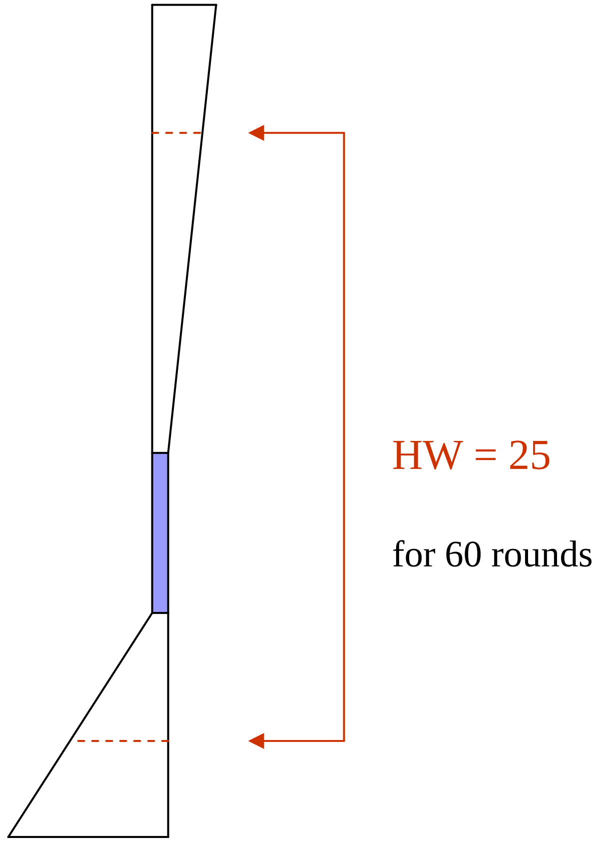
Disturbance vector (DV) for SHA-1



- Change from SHA-0
  - Each x_i is now 32 bits
    - Due to <<1 in message expansion
  - Search space: 2^{16x32}

Search for good DVs for SHA-1

- What are the difficulties?
  - Search space is huge: 2^{512}
  - Hamming weight of DV grows much faster than SHA-0
    - 50+ round seems to be the limit for breaking the 2^{80} barrier
- Main ideas
  - Use heuristic to narrow the search
    - Assume special forms for DV
  - Take advantage of asymmetry in message expansion
  - Remove *all* three conditions on DV
- Matching lower bound
  - HW $\geqslant$ 25 (Jutla, Patthak, last week)



Construct a valid differential path

- What are the difficulties?
  - Local collisions can no longer be stacked together
    - Since all conditions on DV are removed
- How to solve the problem?
  - Derive an impossible path using DV and local collisions
  - Identify un-wanted bit differences
  - Cancel these differences in two ways
    - Carry expansion to introduce a new difference
    - Boolean function to absorb a difference

Derive conditions for differential path

- Conditions on chaining variables
  - Control carry expansion
    - E.g., setting $a_{i,5} = 1$, $a_{i,6} = 0$ expands $\Delta a_{i-1} = 2^5$ to $\Delta a_i = -2^5 + 2^6$
  - Control output difference of $f = (b \wedge c) \vee (\neg b \wedge d)$
    - E.g., setting $c = d$ ensures [$\Delta b = 1 \rightarrow \Delta f = 0$]
- Conditions on message words
  - Set relations among message bits
    - Eliminate carry effect in local collision to increase success prob.
- *Note:* Carry can be good or bad
  - Setting the right conditions can help both ways

Message modification

- Conditions on a_i are of a general form
  - $a_{i,j} = 0, 1$
- Basic idea
  - Round function: $a_i = T + m_{i-1}$
  - Set $a_{i,j}$ = the bit, and *compute* $m_{i-1} = a_i - T$
  - So the condition holds with *p*=1
  - Works when m_i's are independent
- More complex methods: Multi-message modification
  - Use of local collisions

Breaking the 2^{80} barrier

- One-block collision
  - 75-round SHA-1: complexity is less than 2^{80}
    - Already show that “security margin” is not enough
- Near collision
  - 80-round SHA-1: complexity is about 2^{68}
- Two-block collision
  - Use two near collisions
  - Set output differences so that they offset with *probability one*
    - No increase in search complexity
  - Attack complexity is 2^{69}

Summary — cryptanalyst's viewpoint

- Message modification techniques
  - Very effective for boosting success probability
  - Idea applies to any key-less hash function
- Extensive use of local collisions
  - Basic building block for a differential path
  - Also useful in multi-message modification
    - Like a local disturbance without affect global computation
- Manipulation of differential path
  - "Front-loading" path tailored to message modification
  - Turn an impossible path into a possible one
- All techniques leverages on each other

Summary — designer's viewpoint

- The MDx family all follows similar design approaches
  - The M-D iterative structure
    - Some weaknesses found (Joux; Kelsey, Schneier)
  - Message expansion
    - Not enough avalanche effect, even for SHA-1
  - Round function
    - Non-linear components can actually facilitate attack
- What about the SHA-2 family?
  - Local collision existing with smaller prob. (Hawks, Paddon, Rose)
  - Message expansion is much more complicated
  - More analysis is still needed

Thank you very much!