Survey of Cryptanalysis on Hash Functions

Xiaoyun Wang

清華大學

Tsinghua University

Outline

- Design of hash functions
- Earlier cryptanalysis on hash functions
- Recent advances in hash functions cryptanalysis
- SHA-3 competition candidates
- Conclusions

Part I

Design of Hash Functions

Development of Hash Functions

- 1953, IBM discussion
  - Confuse the file keywords
  - Construct the hash table used to computer searching and memory
- 1979，one way hash function, Merkle
  - Hard to find preimage
  - Hard to find second preimage
  - Guarantee secure authentication serve

Cryptographic Hash Function

- Davies, Price， hash functions used to digital signatures, Technical Report， 1980
  - Destroy the algebraic structure of RSA signature to resist on the existential forgery attack:

    $$S(M_1M_2)=S(M_1)\ S(M_2)$$

  - Improve the signature efficiency
- Signature of message M is computed as:

$$s=S(h(M))$$

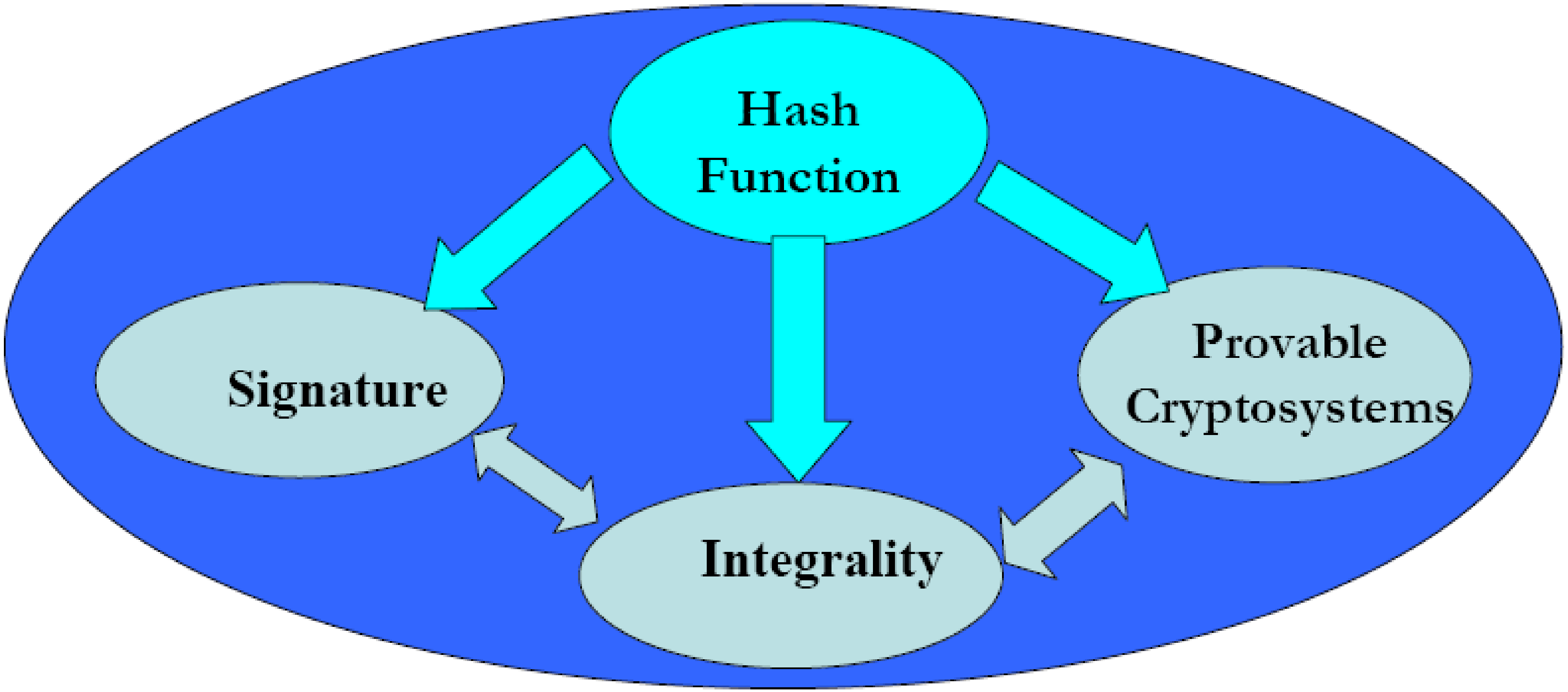
h is the hash function

Hash Function is One of Fundamental Cryptographic Algorithms

- One of three fundamental cryptographic algorithms
- Three fundamental cryptographic algorithms: encryption, signature, hash function
- Widely used in the security of network and wireless communication

Hash Function is One of Fundamental Cryptographic Algorithms

- For example, hash function is the key technique to design bit commitment



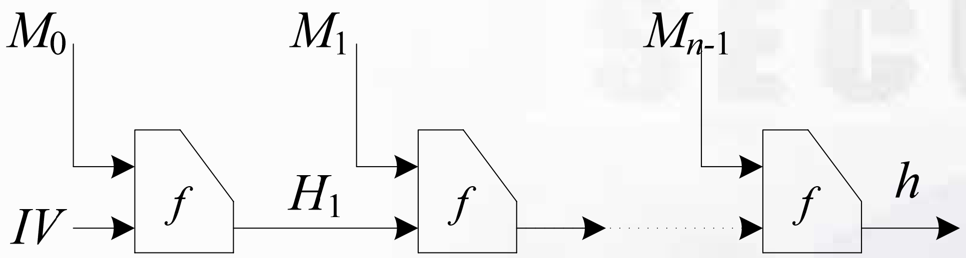
Design Principle of Hash Functions

- Merkle-Damgård Meta Method, Crypto 89
  - Given message with padding $M=(M_0, M_1, ..., M_{n-1})$， the hash value of M is computed as

$$H_0=IV$$

$$H_i=f(H_{i-1}, M_{i-1}),\quad 0<i<n+1$$

$$h = H_n$$



Hash Functions Based on Block Ciphers

- Hash function with one-block length
- Secure hash functions, concluded by Preneel, 1993

Matyas-Meyer-Oseas

$$H_i = E_{H_{i-1}}(M_i) \oplus M_i \qquad H_i = E_{M_i}(H_{i-1}) \oplus M_i \oplus H_{i-1}$$

$$H_i = E_{H_{i-1}}(M_i \oplus H_{i-1}) \oplus M_i \oplus H_{i-1} \qquad H_i = E_{M_i}(M_i \oplus H_{i-1}) \oplus H_{i-1}$$

Miyaguchi-Preneel

$$H_i = E_{H_{i-1}}(M_i) \oplus H_{i-1} \oplus M_i \qquad H_i = E_{M_i \oplus H_{i-1}}(M_i) \oplus M_i$$

$$H_i = E_{H_{i-1}}(M_i \oplus H_{i-1}) \oplus M_i \qquad H_i = E_{M_i \oplus H_{i-1}}(H_{i-1}) \oplus H_{i-1}$$

Davies-Meyer

$$H_i = E_{M_i}(H_{i-1}) \oplus H_{i-1} \qquad H_i = E_{M_i \oplus H_{i-1}}(M_i) \oplus H_{i-1}$$

$$H_i = E_{M_i}(M_i \oplus H_{i-1}) \oplus M_i \oplus H_{i-1} \qquad H_i = E_{M_i \oplus H_{i-1}}(H_{i-1}) \oplus M_i$$

Hash Functions Based on Block Ciphers

- Hash function with double(multi)-block length
  - MDC-2, MDC-4,1990，Brachtl etc,
    (MDC-2 ANSI X9.31 standard)
  - Parallel Davies-Meyer，Lai, Massey, Eurocrypt 92
  - GOST, Russia stardard
  - ……..

Dedicated Hash Functions

- MDx family: proposed by Rivest
  - MD4, Crypt 90
  - MD5, RFC 1992
- SHA family: proposed by NIST
  - SHA-0, FIPS-180, 1993
  - SHA-1, FIPS-180-1, 1995
  - SHA-2 (SHA-256/384/512), FIPS-180-2, 2002

Dedicated Hash Functions

- RIPEMD family
  - RIPEMD: RIPE project, 1995
  - RIPEMD-160: Dobbertin, Bosselaers, Preneel, 1996
- Some other hash functions
  - HAVAL, Tiger, Whirpool etc

Part II

Earlier Cryptanalysis on Hash Functions

Earlier Cryptanalysis on Hash Functions Based on Block Ciphers

- Mainly focus on the structure attack
- Many hash functions based on block ciphers are broken by Preneel et al., PH. D thesis, 2003
- The 12 secure structures are listed by Preneel: strong secure 8

Earlier Cryptanalysis on Dedicated Hash Functions

- Collision attack on MD4, Dobbertin, FSE 1996
  - Find a collision on MD4 with probability 2^{-22}
  - Differential attack and mathematical equations
- Not one way for 2-round MD4, Dobbertin, FSE 1998
- Not collision-free for 2-round RIPEMD, J. of Cryptology, 1998

Earlier Cryptanalysis on Dedicated Hash Functions

- Free-start collision of MD5, Boer and Bosselaers, Eurocrypto’93
  - Same message with two different initial values
  - Weak avalanche for the most significant bit
  - The differential path with high probability is successfully used to analyzing MACs based on MD5 (in 2005-2006 and 2009)
- Semi free-start collision of MD5, Dobbertin, Eurocrypt’96 Rump Session
  - Two different 512-bit messages with a chosen initial value

Earlier Cryptanalysis on Dedicated Hash Functions

- SHA-0 differential attack, Chabaud, Joux, Crypto'98
  - Two collision differential paths are found, and each path can be divided into 6-step local collisions
- Another SHA-0 attack in 1997 (Wang, in Chinese, not published)
  - Same collision paths by solving mathematical equations:

    2 solutions of 2^{512} message difference space
  - The theoretic support for SHA-1 cryptanalysis

Part III

Recent Advances in Hash Functions Cryptanalysis

Recent Collision Attack on Hash Functions (I)

Bit Carry

Mathematical Characteristic

Bit Tracing

Muiti-Block Collision

Message Avalanch Control
Message Modification

Convert Impossible differntial to possible differntial

~~MD4~~ ~~RIPEMD~~ ~~HAVAL~~

~~MD5~~ ~~SHA-0~~ ~~SHA-1~~

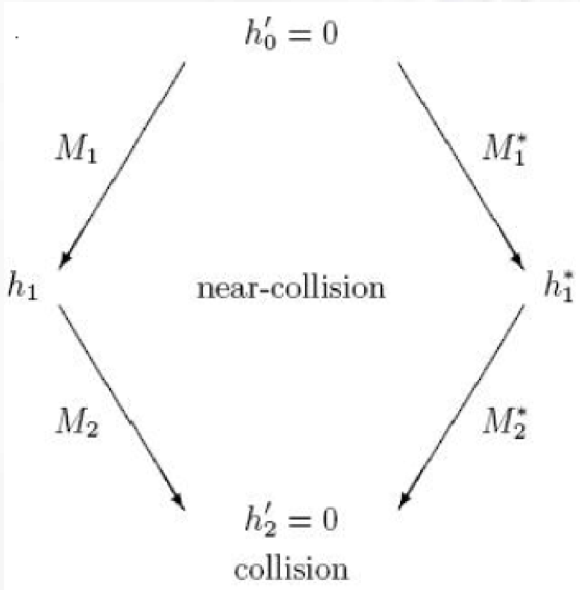
Recent Collision Attack on Hash Functions (I)

		m_i	s	Δm_i								
2	d_1	m_1	7	2^{31}	7							
3	c_1	m_2	11	$-2^{28}+2^{31}$			11			8		
4	b_1	m_3	19			26						
5	a_2	m_4	3									
6	d_2	m_5	7		14							
7	c_2	m_6	11				22			19,20, -21		
8	b_2	m_7	19			-13, -14, 15						
9	a_3	m_8	3		17							
10	d_3	m_9	7					20, -21, -22, 23				-26
11	c_3	m_{10}	11							-30		
12	b_3	m_{11}	19			32						
13	a_3	m_{12}	3	-2^{16}	-17*				23		26	
14	d_3	m_{13}	7					-27, -29, 30				
15	c_3	m_{14}	11									
16	b_3	m_{15}	19			19						
17	a_4	m_0	3						-26, 27		-29, -32	
18	d_4	m_4	5									
19	c_4	m_8	9									
20	b_4	m_{12}	13	-2^{16}		32, -30						
21	a_5	m_1	3	2^{31}					-29, 30	32*	-32	
22	d_5	m_5	5									
23	c_5	m_9	9									
24	b_5	m_{13}	13									
25	a_6	m_2	3	$-2^{28}+2^{31}$					-29*		32*	
26	d_6	m_6	5									
36	b_9	m_{12}	15	-2^{16}	32							
37	a_{10}	m_2	3	$-2^{28}+2^{31}$	32							
38	d_{10}	m_{10}	9									
39	c_{10}	m_6	11									
40	b_{10}	m_{14}	15									
41	a_{11}	m_1	3	2^{31}	32*							

Bit tracing to find the collision path for MD4

Recent Collision Attack on Hash Functions (I)

- Multi-block collision, Joux etc, Crypto 04 Rump Session, Formalized by Biham and Joux etc in Eurocrypt 05
- Independently proposed collision attack with two message blocks for MD5, Wang and Yu at Crypto 04 Rump Session



Collision Attacks and Practical Attacks (II)

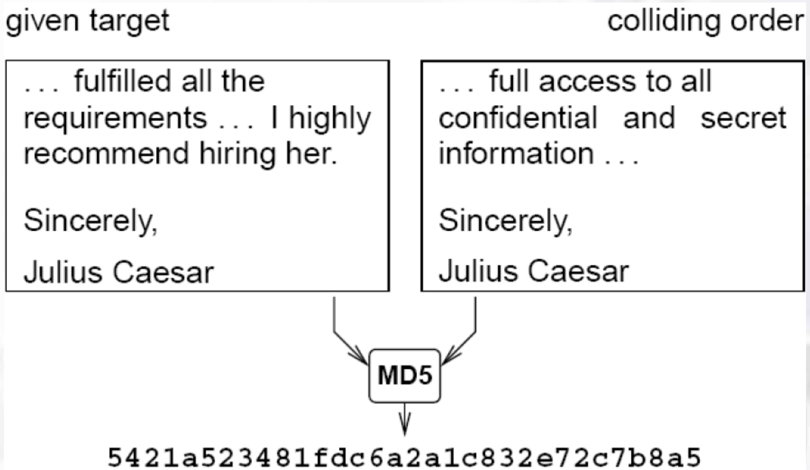
- PS editor files with same signature, Lucks and Daum, Rump Session in Eurocrypt'05
  - R_1 and R_2 is a random collision pair
  - Editor software with redundancy

The target documents are T_1 and T_2:

$$Y_1 = \overbrace{\text{preamble; } \mathtt{put}(R_1);}^{X_1}\ \underbrace{\mathtt{put}(R_1);\ \mathtt{if}(=)\ \mathtt{then}\ T_1\ \mathtt{else}\ T_2;}_{S}$$

$$Y_2 = \underbrace{\text{preamble; } \mathtt{put}(R_2);}_{X_2}\ \overbrace{\mathtt{put}(R_1);\ \mathtt{if}(=)\ \mathtt{then}\ T_1\ \mathtt{else}\ T_2;}^{S}$$

- Viewing Y_1: $R_1 = R_1$, thus T_1 is displayed.
- Viewing Y_2: $R_2 \neq R_1$, thus T_2 is displayed.



- Other editor softwares PDF,TIFF and Word 97, Gebhardt et.al, NIST Hash Function Workshop 2005

Collision Attacks and Practical Attacks (II)

- Colliding valid X.509 certificates
  - Lenstra, Wang, Weger, forged X.509 certificates，http://eprint.iacr.org/2005/067.pdf

    Same owner with different public keys (2048 bits)
  - Stevens, Lenstra, Weger, Eurocrypt 2007

    8192-bit public key (8-block collision)
  - Stevens etc, Crypto 2009

    Pass the browser authentication, different owners, different public keys

  US-CERT：MD5 vulnerable to collision attacks

Preimage Attacks on Hash Functions (III)

- Preimage attacks on hash functions, Leurent, FSE 2008
  - Partial pseudo preimage attack on the compression function of MD4. Choose 64-bit of the output for the cost of 2^{32} compression function computations
  - Preimage attack on compression function of MD4 with complexity 2^{96}
  - Attack on the full MD4 with complexity 2^{102} using birthday paradox and layered hash tree

Preimage Attacks on Hash Functions (III)

- Aoki and Sasaki, preimage attacks on one-block MD4, 63-step MD5, SAC 2009
  - A preimage of one-block MD4 can be found with 2^{107} MD4 computations
  - A preimage of MD5 reduced to 63 steps can be found with 2^{121} MD5 computations
- Sasaki and Aoki, preimage attack on full MD5, Eurocrypt 2009
  - Searches a pseudo-preimage with complexity $2^{116.9}$
  - Searches a preimage with complexity $2^{123.4}$

Collision Attacks and MAC Cryptanalysis (**IV**)

- Key recovery of envelop MAC based on MD4, Yu and Wang, Ecrypt hash function workshop 2005
- Contini,Yin, Asiacrypt 2006
  - Partial key recovery attacks on HMAC/NMAC-MD4/SHA-0

Collision Attacks and MAC Cryptanalysis (**IV**)

- Fouque, Leurent, Nguyen, Crypto 2007
  - Full key recovery attack on HMAC/NMAC-MD4
  - Full key recovery attack on NMAC-MD5 in the related-key setting
- Wang, Ohta, Kunihiro, Eurocrypt 2008
  - Improved outer-key recovery attacks on HMAC/NMAC-MD4
  - Improved outer-key recovery attacks on NMAC-MD5 in the related-key setting

Collision Attacks and MAC Cryptanalysis (**IV**)

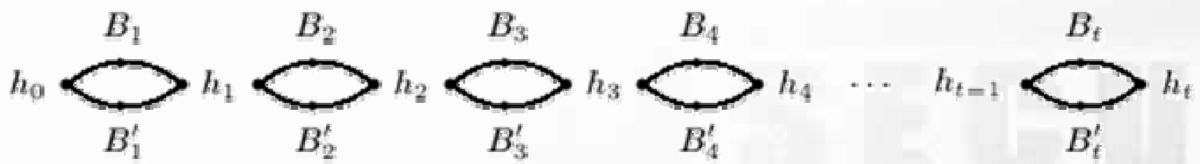
- Distinguishing-H attack on MAC/NMAC-MD5, MD5-MAC, Eurocrypt 09
  - New birthday attack to detect the collision (near-collision) with differential path instead of only collision detection
  - Partial key recovery attack on MD5-MAC
- The birthday Distinguishing-R attack for all the iterated MACs, Preneel and van Oorschot，Crypto'95

Cryptanalysis of MD Structure (**V**)

- Length extension attack (fast implementation)
  - Given $h = H(M)$, M is unknown, by choosing M', an adversary can calculate:

$$h' = H(M \| M') = H(h, M')$$

  - If $H(M) = H(N)$， then $H(M \| S) = H(N \| S)$
- Multi-collision attack： $t2^{n/2}$, ideal complexity： $2^{\frac{t-1}{t}\cdot n}$



- Fixed point attack：

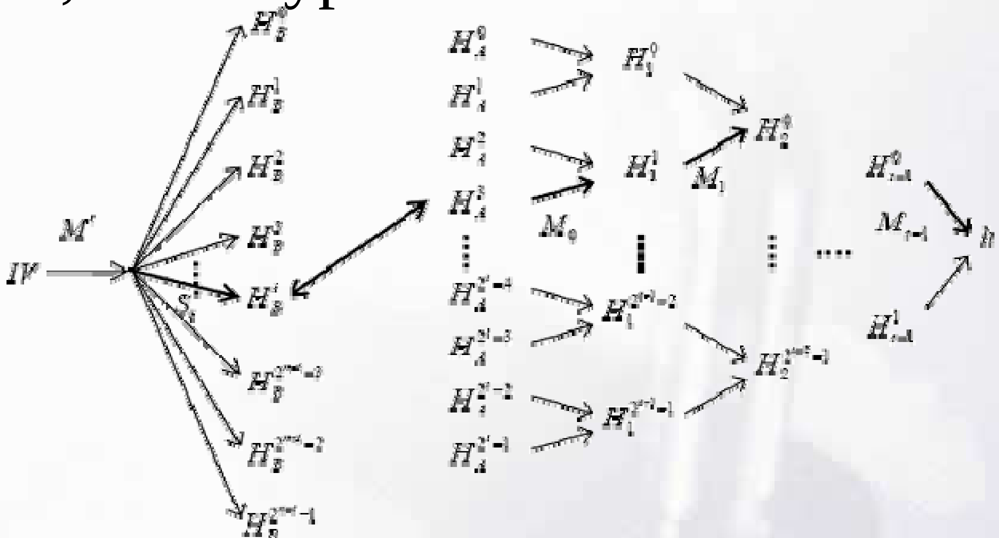
$$f(h_p, M) = h_p$$

Cryptanalysis of MD Structure (**V**)

- Kelsey, Schneier, Second preimage attack of long messages, Eurocrypt 2005
- Second preimage attack based on fixed points
  - Complexity： $\max\{\ 2^{n-k},\ 2^{k}\}$
  - Message length: 2^k bits
- Second preimage attack based on Joux's multicollisions
  - Complexity: $k\cdot 2^{\frac{n}{2}+1}+2^{n-k+1}$

Cryptanalysis of MD Structure (**V**)

- Kelsey, Kohn, Herding attack, Eurocrypt 2006



- Details of the attack
  - Choose messages (important or not) M=(M_0, …..,$M_{t\text{-}1}$) with different IVs to produce h=H(M) by birthday attack
  - Choose $2^{n/2}$ important or sensitive message M'
  - Search M' and M such that h=H(M'||M) by birthday attack
- Complexity：$2^{t/2+n/2+2}+2^{n-t}+2^{n-k}$

Rebounded Attack on Hash Functions (**VI**)

- Introduced by Mendel et al., FSE 2009
  - If there is a truncated differential path of half rounds less than half of birthday complexity, the attack works
  - Reduced Whirlpool and Grϕstl, FSE 2009
  - Rebounded attack on the full lane compression function, Asiacrypt 2009
  - …….

Part IV

SHA-3 Competition Candidates

Security Requirements of the Hash Functions

- Collision resistance of approximately *n/2* bits ($2^{n/2}$ computations)
- Preimage resistance of approximately *n* bits
- Second-preimage resistance of approximately *n-k* bits for any message shorter than 2^k *bits* (for MD construction)
- Resistance to length-extension attacks (usually MD construction is prohibited)
- Truncating *m*-bit of the candidate function's output, the security parameter is *m* replacing *n*

Notes on the Security Requirements

- Resistance to length-extension attacks
  - Resistance to multi-block collision attacks
  - Resistance to multi-collision attacks
  - Resistance to second preimage attacks of long messages and herding attack
- Second preimage resistance of approximately *n* bits for messages with any length (strong requirement)
  - Security requirements for non-MD constructions

First Round Candidates

- 2008.10.31, NIST received 64 algorithms
  - AES project received 21 algorithms
  - More attention to hash functions
- 2008.12.10：51 algorithms satisfy the Minimum Acceptability Requirements

Second Round Candidates

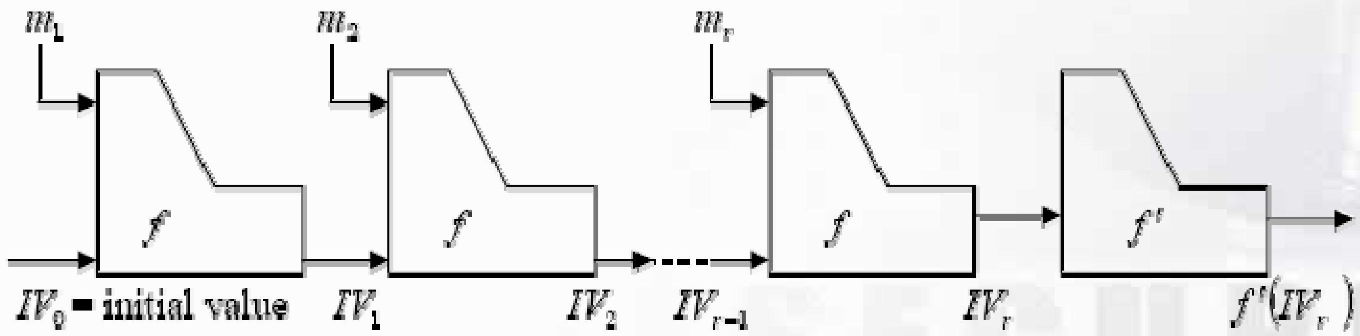
- 5 Sponges，2 HAIFAs，5 Wide Pipes，1 Wide Pipe HAIFA

Algorithm	Structure
BLAKE	HAIFA
BMW	Wide Pipe
CubeHash	Sponge
ECHO	Wide Pipe, HAIFA
Fugue	Sponge
Grostl	Wide Pipe
Hamsi	Sponge

Algorithm	Structure
JH	Wide Pipe
Keccak	Sponge
Luffa	Sponge
Shabal	Wide Pipe
SHAvite-3	HAIFA
SIMD	Wide Pipe
Skein	UBI chaining

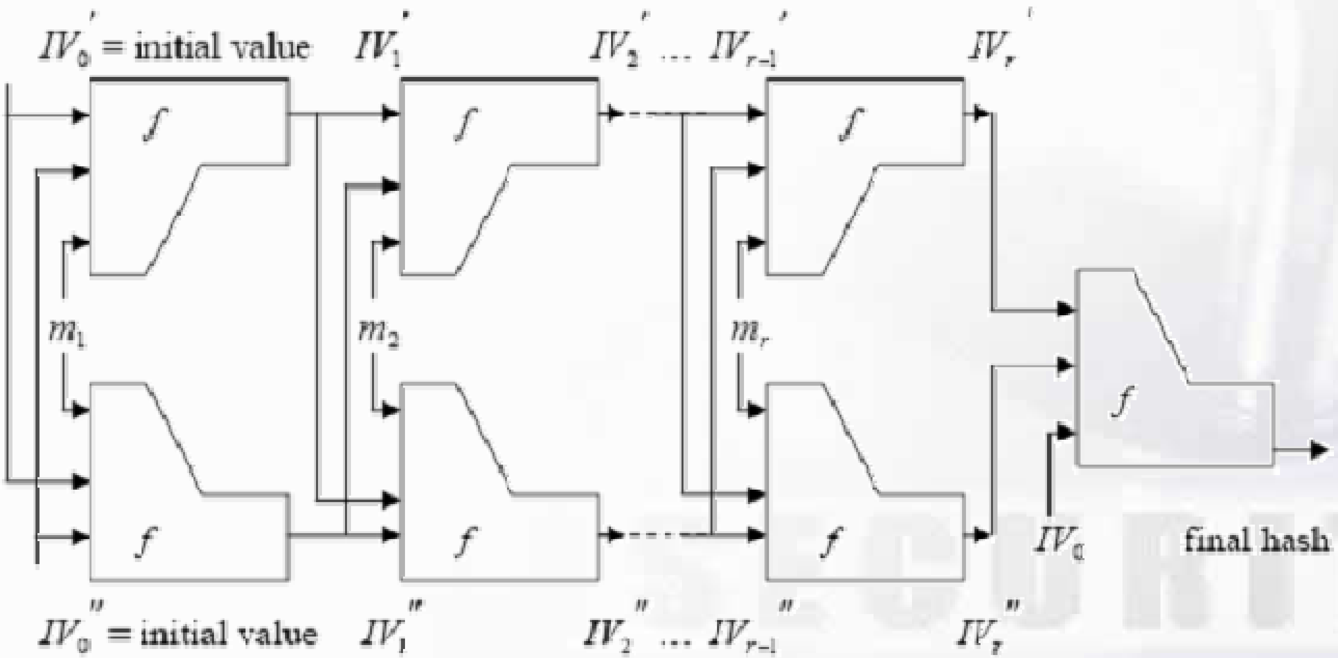
Main Structures of SHA-3 Candidates

- Wide Pipe, Lucks, Asiacrypt 2005
- Compress function： $f:\{0,1\}^w \times \{0,1\}^p \rightarrow \{0,1\}^w$
- Truncation function： $f':\{0,1\}^w \rightarrow \{0,1\}^n$



Main Structures of SHA-3 Candidates

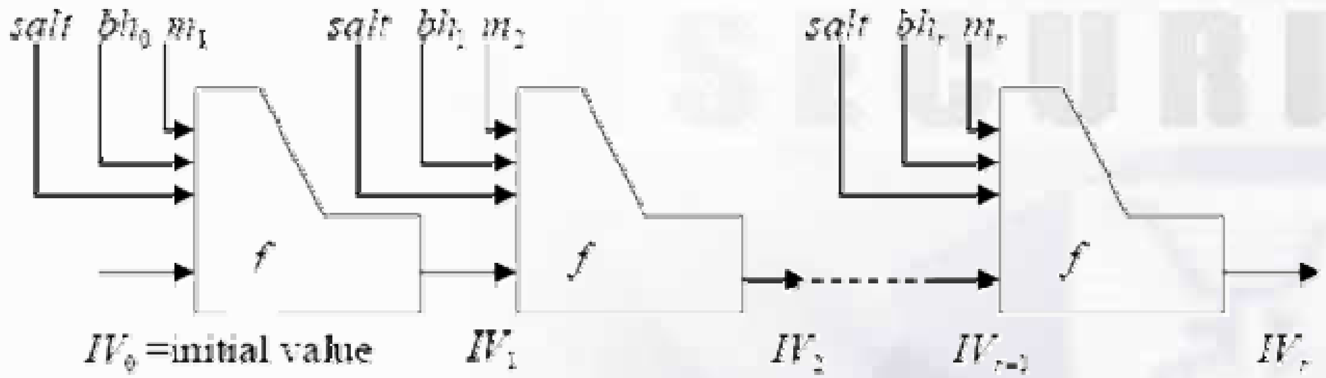
- Double Pipe, Lucks, Asiacrypt 2005



Main Structures of SHA-3 Candidates

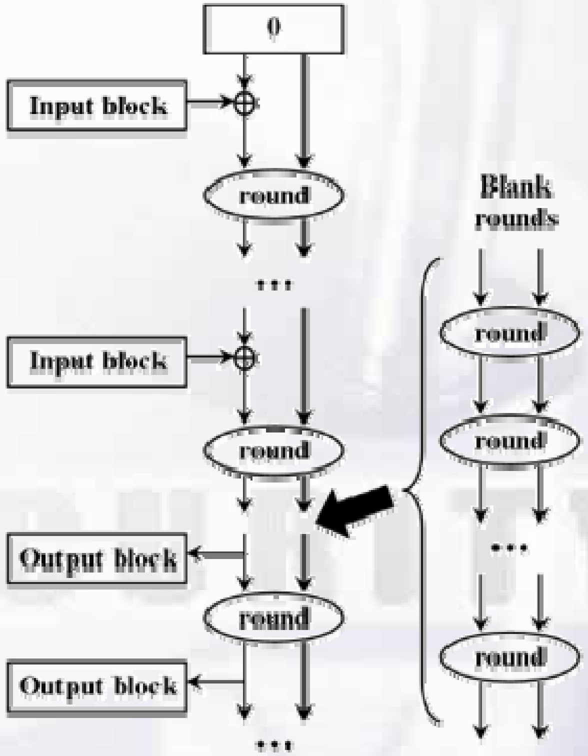
- HAIFA , Biham etc., Cryptographic Hash WorkShop, 2006
- Salt+bh_i： n/2 bits， the ideal strength for computing second preimage seems to be $2^{n/2+n/2}$
- Computational efficiency is (m-n/2)/m times that of MD structure, where n is the output length and m is the message block size

e.g. the output length is 256 bits, message block size is 512 bits, then the efficiency is (512-128)/512=0.75 times



Main Structures of SHA-3 Candidates

- Sponge, Bertoni etc.， Ecrypt workshop on hash functions, 2007
- Provable security
  - If each iteration is secure
- Building block is a reduced block cipher PANAMA ，RADIOGATúN etc
- Building block is a full block cipher



Security Status of First Round SHA-3 Candidates

Hash Name	Principal Submitter	Best Attack on Main NIST Requirements	Best Attack on other Hash Requirements
ARIRANG	Jongin Lim		near-collision
AURORA	Masahiro Fujita	2nd preimage	
Blender	Colin Bradbury	collision, preimage	near-collision
Cheetah	Dmitry Khovratovich		length-extension
CHI	Phillip Hawkes		pseudo-2nd preimage
CRUNCH	Jacques Patarin		length-extension
Dynamic SHA	Xu Zijie	collision	length-extension
Dynamic SHA2	Xu Zijie	collision	length-extension
ECOH	Daniel R. L. Brown	2nd preimage	
Edon-R	Danilo Gligoroski	preimage	

Note：from SHA-3 ZOO

Security Status of First Round SHA-3 Candidates

EnRUPT	Sean O'Neil	collision	
ESSENCE	Jason Worth Martin	collision	
FSB	Matthieu Finiasz		
LANE	Sebastiaan Indesteege		semi-free-start collision
Lesamnta	Hirotaka Yoshida		pseudo-collision
LUX	Ivica Nikolić	collision, 2nd preimage	DRBG,HMAC
MCSSHA-3	Mikhail Maslennikov	2nd preimage	
MD6	Ronald L. Rivest		
NaSHA	Smile Markovski	collision	

Note： from SHA-3 ZOO

Security Status of First Round SHA-3 Candidates

SANDstorm	Rich Schroeppel		
Sarmal	Kerem Varıcı	preimage	
Sgàil	Peter Maxwell	collision	
Spectral Hash	Çetin Kaya Koç	collision	
SWIFFTX	Daniele Micciancio		
TIB3	Daniel Penazzi	collision	
Twister	Michael Gorski	preimage	
Vortex	Michael Kounavis	preimage	

Note： from SHA-3 ZOO

Security Status of Second Round SHA-3 Candidates

Algorithms	Cryptanalytic Results	Com-plexity	Authors
Blake	4(out of 10) rounds near-collision of Blake-256	2^{42}	Guo etc.
	5(out of 10) rounds impossible differential of Blake-224/256 6(out of 14) rounds impossible differential of Blake-384/512	— —	Aumasson etc.
ECHO	7(out of 8) rounds distinguisher of ECHO-224/256 7(out of 10) rounds distinguisher of ECHO-384/512	2^{384} 2^{384}	Mendel etc.
JH	pseudo-collision pseudo-2nd preimage	— —	Bagheri
Keccak	16(out of 24) rounds distinguisher	$2^{1203.88}$	Aumasson etc.
	18 rounds distinguisher	2^{1370}	Boura etc.
CubeHash r/b r: rounds b: block size(byte)	preimage attack	2^{513-4b}	Aumasson etc.
	second preimage attack onCubeHash 6/4 collision attack on CubeHash 6/16	2^{478} 2^{222}	Brier etc.

Security Status of Second Round SHA-3 Candidates

Algo-rithms	Cryptanalytic Results	Com-plexity	Authors
Grøstl	6 (out of 10) rounds semi-free-start collision of Grøstl-256 7 rounds distinguisher of the permutation of Grøstl-256 7 rounds distinguisher of the output transformation of Grøstl-256	2^{64} 2^{55} 2^{56}	Mendel etc.
SHAv-ite-3	example for chosen-salt, chosen-counter pseudo-collision	—	Peyrin
	fixed points on SHAvite-3-256 block cipher	—	Nandi
Shabal	non-randomness	—	Knudsen etc.
	non-randomness	—	Aumasson etc.
BMW	example of near-collision(original version) pseudo-preimage(original version) pseudo-collision(original version)	$2^{3n/8+1}$ $2^{3n/4+1}$	Thomsen
Skein	17 rounds(out of 72) pseudo near-collision on Skein-512(original version) 35 rounds known related-key distinguisher of Threefish-512(original version) 32 rounds related-key attack onThreefish-512 (original version)	2^{24} 2^{478} 2^{312}	Aumasson etc.

Security Status of Second Round SHA-3 Candidates

Algo-rithms	Cryptanalytic Results	Com-plexity	Authors
Hamsi	non-randomness of 5 rounds(out of 3/6) Hamsi-224/256 6 rounds distinguisher of Hamsi-224/256 12 rounds(out of 6/12) distinguisher of Hamsi-384/512	2^{27} 2^{729}	Aumasson etc.
	3 rounds pseudo near-collision of Hamsi-256	2^{21}	Nikolić
	3 rounds pseudo near-collision of Hamsi-256 4 rounds differential path of Hamsi-256 5 rounds differential path of Hamsi-256	2^{5} 2^{32} 2^{125}	Wang etc.
Luffa	zero-sum distinguisher on Q permutation	2^{82}	Aumasson etc.
	examples of pseudo collision, pseudo second preimage example of pseudo preimage of Luffa-256 pseudo preimage attack on Luffa-384/512 differential paths of *Q* permutation	$2^{64}/2^{128}$ 2^{214}	Jia etc.

Conclusions

- Today, it is more clear with collision attack, second pre-image attack, preimage attack and their relationship on the existing dedicated hash functions
- More clear with influence of hash cryptanalysis on MACs cryptanalysis
- More clear with the design of hash function structures, and compression functions

Thanks!