Lecture 7

HASH FUNCTIONS

R. Rhouma

UTAS
Sultanate of Oman
September 2022

CSSY2201 : Introduction to Cryptography

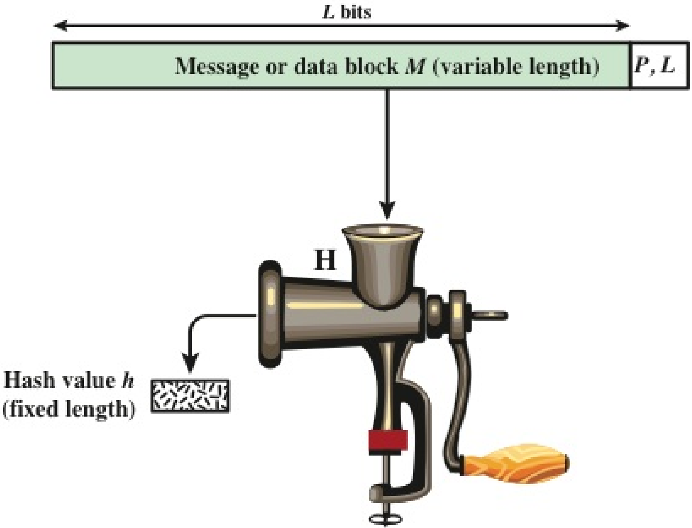
Plan

1. Hash functions
2. Secure hash Algorithm : SHA

Hash functions

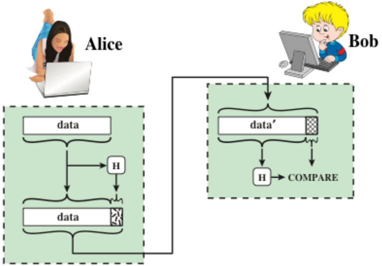
- Hash function accepts variable-length input and outputs a digest or hash of fixed length
- $h = H(M)$
- Its main purpose is integrity checking
- A cryptographic hash function is an algorithm that is mathematically difficult to :
  - Find an entry that gives a well-specified digest (Property : one-way function)
  - find two entries that give the same digest (Property : Collision-free)

Cryptographic Hash function h=H(M)

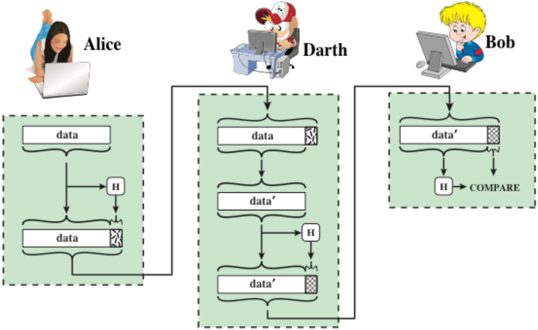


the output of a hash function from a message is named digest

- The digest of the message is its fingerprint
- Much smaller than original post
- easy to calculate
- cannot find message from digest
- Changing the message automatically changes the digest



(a) Use of hash function to check data integrity

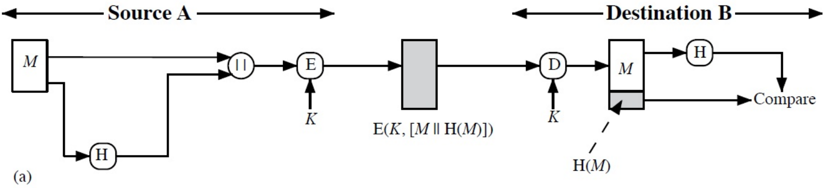


(b) Man-in-the-middle attack

جامعة التقنية والعلوم التطبيقية
University of Technology and Applied Sciences
Salalah صلالة

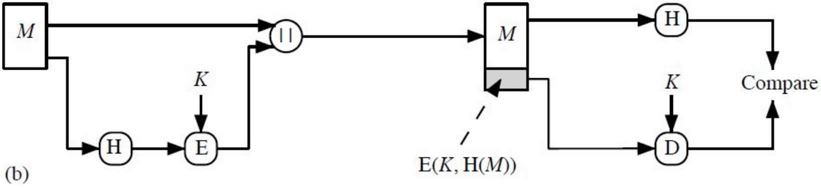
Message Authentication

- Verify message integrity
  - Ensure received data is exactly as sent
  - Ensure sender identity is valid
- Example 1 : Encrypt the message and its digest with a symmetric cryptosystem



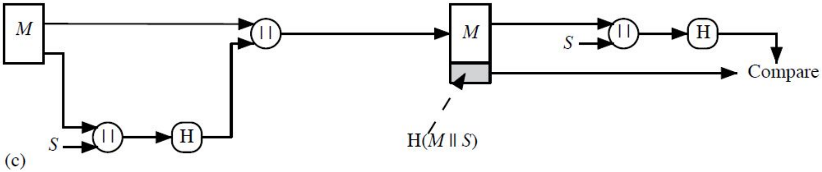
Message Authentication

- Example 2 : Encrypt only the message digest
- it reduces the complexity of calculation if confidentiality is not requested



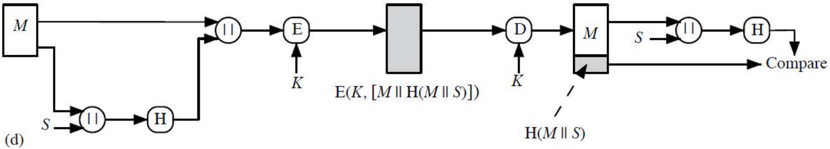
Message Authentication

- Example 3 : A shared secret is hashed
- No need for encryption



Message Authentication

- Example 4 : A shared secret combined with confidentiality



Other Uses of Hash Functions

- Used to create password files
  - When a user types a password, the hash of the password is compared to the saved hash for verification
  - This approach is used by the majority of operating systems
- used to detect intrusions and viruses
  - save the H(f) of each file to disk
  - the antivirus can later check if the file has been altered or not by recalculating its digest H(f)
  - An intruder will try to change F without changing H(f) : very difficult !
- can be used to build PRNG pseudo-random sequence generators
  - generate keystreams, secret keys

Requirements of a hash function

- Variable length entry
- Fixed length output
- Efficiency : given x, it is easy to generate the digest H(x) in s/w or h/w
- One-way function (Pre-image resistant) : For a given digest h, it is impossible to find y such that $H(y) = h$
- No broad sense collision (Second pre-image resistant : weak collision resistant) : For any given x, it is impossible to find $y \neq x$ such that $H(y) = H(x)$
- No collision in the strict sense (collision resistant : Strong collision resistant) : it is impossible to find a pair (x, y) such that $H(x) = H(y)$
- Random criterion : The output of H must be random according to the standard tests (NIST : 16 tests of the random criterion)

NB : "impossible" = "mathematically or by calculation difficult"

Requirements of a hash function

	Preimage Resistant	Second Preimage Resistant	Collision Resistant
Hash + digital signature	yes	yes	yes*
Intrusion detection and virus detection		yes	
Hash + symmetric encryption			
One-way password file	yes		
MAC	yes	yes	yes*

* Resistance required if attacker is able to mount a chosen message attack

Birthday Paradox

- In a class, what is the probability that 2 students celebrate their birthdays on the same day ?
- With 365 days a year, about thirty students in the class, we say to ourselves that it must be weak...
- We will calculate the probability that, in a group of k people, these people all have a different birthday :
  - If there are 2 people, the first can have a birthday anytime, the second any other day. we therefore have : $p_2 = \frac{364}{365} = 1 - \frac{1}{365}$
  - if mnt we have k people : $p_3 = (1 - \frac{1}{365})(1 - \frac{2}{365})$
  - in a group of k people, $p_k = (1 - \frac{1}{365})(1 - \frac{2}{365}) \ldots (1 - \frac{k-1}{365})$

Student Numbers	Prob. that their **anniv. dont match**
1	1
2	0.99
5	0.97
10	0.88
20	0.58
22	0.52
23	0.49
30	0.29
50	0.03

$\Longrightarrow$ It therefore only takes 23 people for there to be more than one chance in 2 (chance > 0.5) for 2 people to have their birthday on the same day.

Attacks based on the Birthday Paradox

- In an attack that looks for collisions, the adversary wants to find 2 messages that give the same digest.
- In a class of 23 students the probability of finding 2 students with the same birthday is > 0.5
- If the digest is encoded on b bits, there are 2^b possible hashes.
- if we take k different msgs, the probability of finding 2 msgs with the same digest is :

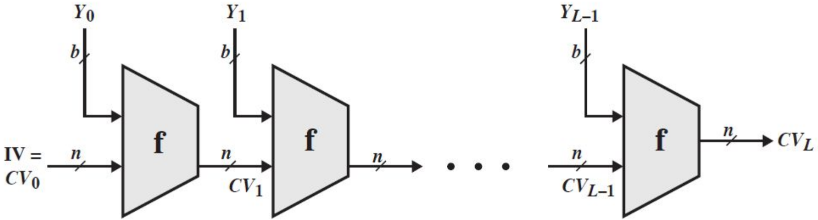
$$p = 1 - (1 - \frac{1}{2^b})(1 - \frac{2}{2^b})\ldots(1 - \frac{k-1}{2^b})$$

- for $p \geq \frac{1}{2}$, it suffices that $\overline{p} = (1 - \frac{1}{2^b})(1 - \frac{2}{2^b})\ldots(1 - \frac{k-1}{2^b}) \leq \frac{1}{2}$
- we have $(1 - \frac{j}{2^b}) \sim e^{(-\frac{j}{2^b})}$
- we then have $(1 - \frac{1}{2^b})(1 - \frac{2}{2^b})\ldots(1 - \frac{k-1}{2^b}) \sim e^{(-\frac{k(k-1)}{2^{b+1}})}$
- it is therefore necessary that $e^{(-\frac{k(k-1)}{2^{b+1}})} \leq \frac{1}{2}$

Digest Size	Total nb of trials to find a collision
8	20
16	302
32	77169
64	2 x 10^{10}
128	9 x 10^{19}
160	8 x 10^{25}
256	3 x 10^{40}

جامعة التقنية والعلوم التطبيقية
University of Technology and Applied Sciences
Salalah صلالة

General structure of a hash function



IV	=	Initial value	L	=	number of input blocks
CV_i	=	chaining variable	n	=	length of hash code
Y_i	=	ith input block	b	=	length of input block
f	=	compression algorithm			

Secure Hash Algorithm (SHA)

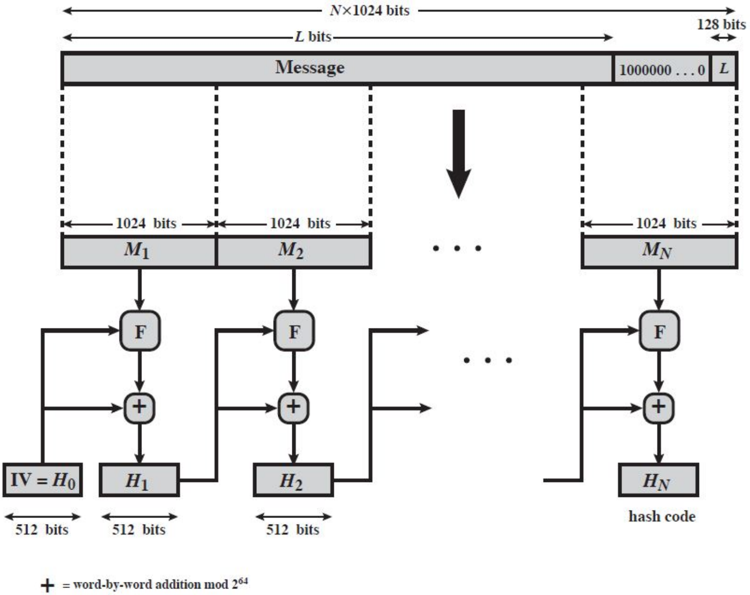
- SHA was designed by "National Institute of Standards and Technology (NIST)" and published as "federal information processing standard" (FIPS 180) in 1993
- was revised in 1995 as SHA-1
- Based on MD4 hash function
- Produces a 160-bit size digest
- In 2002 NIST produced a revised version of the standard to define 3 more SHAs with lengths 256, 384, and 512 Known as SHA-2

Comparaison of SHA versions

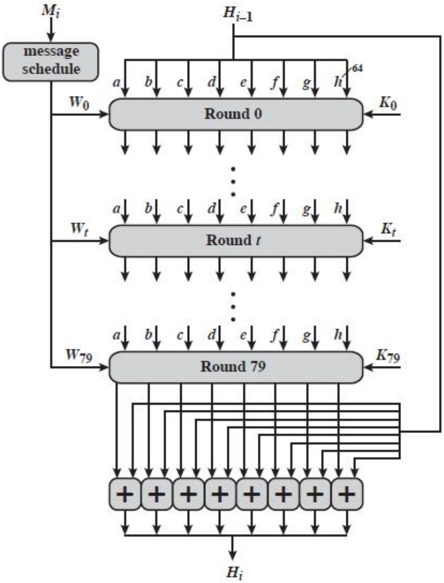
	SHA-1	SHA-224	SHA-256	SHA-384	SHA-512
Message Digest Size	160	224	256	384	512
Message Size	$< 2^{64}$	$< 2^{64}$	$< 2^{64}$	$< 2^{128}$	$< 2^{128}$
Block Size	512	512	512	1024	1024
Word Size	32	32	32	64	64
Number of Steps	80	64	64	80	80

جامعة التقنية والعلوم التطبيقية
University of Technology and Applied Sciences
Salalah صلالة

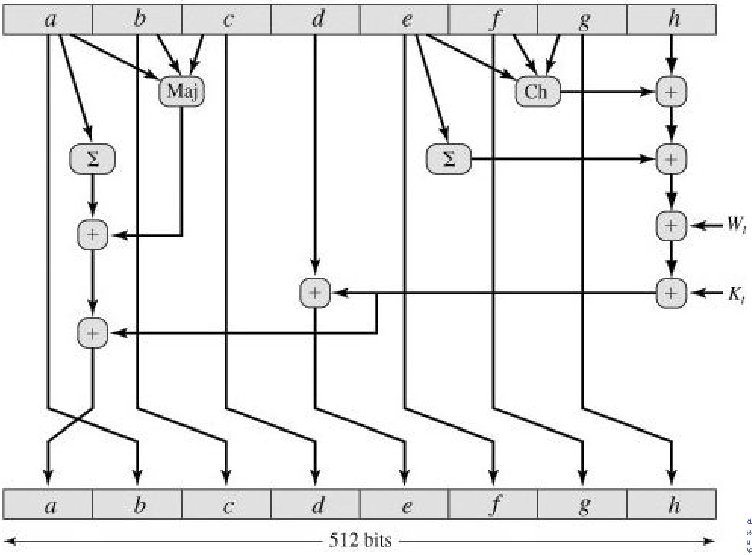
SHA-512


N×1024 bits
L bits
128 bits
Message
1000000 . . . 0
L
1024 bits
1024 bits
1024 bits
M1
M2
MN
F
F
F
IV = H0
H1
H2
HN
512 bits
512 bits
512 bits
hash code
+ = word-by-word addition mod 2^64

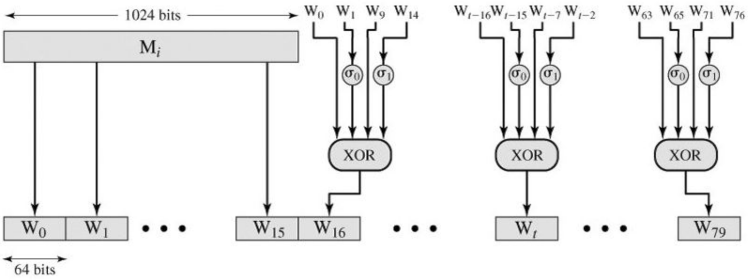
SHA-512 : Processing of a 1024-Bit block



SHA-512 : buffers update



SHA-512 : Processing of the message M_i



جامعة التقنية
والعلوم التطبيقية
University of Technology
and Applied Sciences
Salalah صلالة