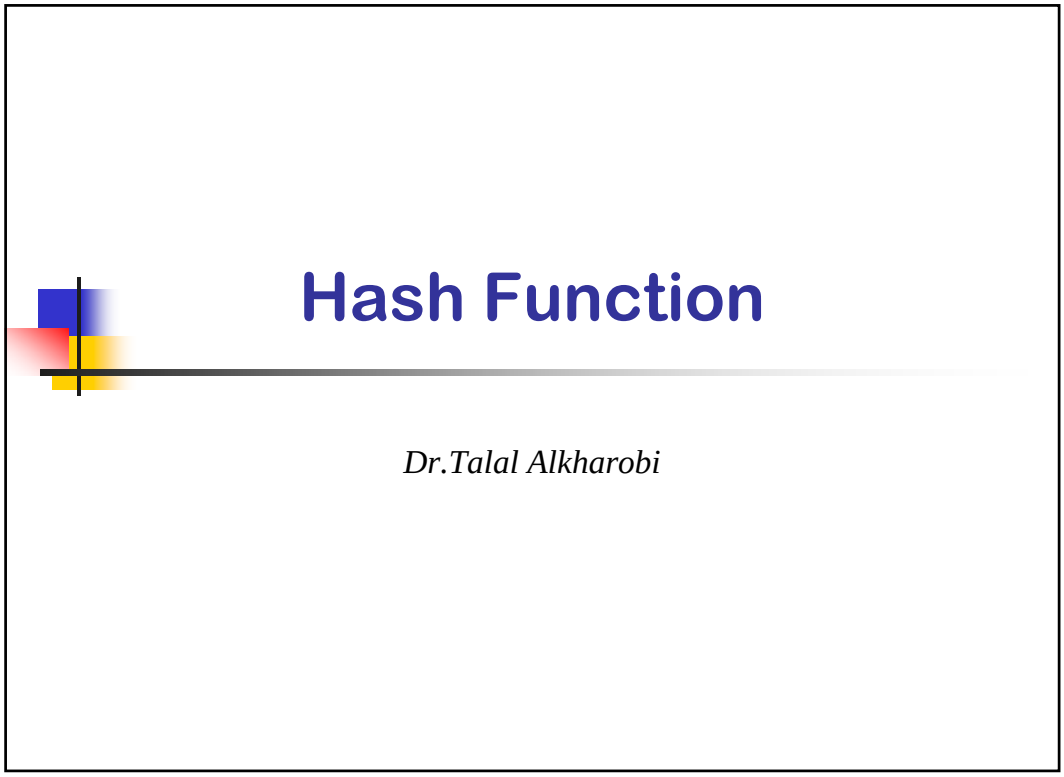
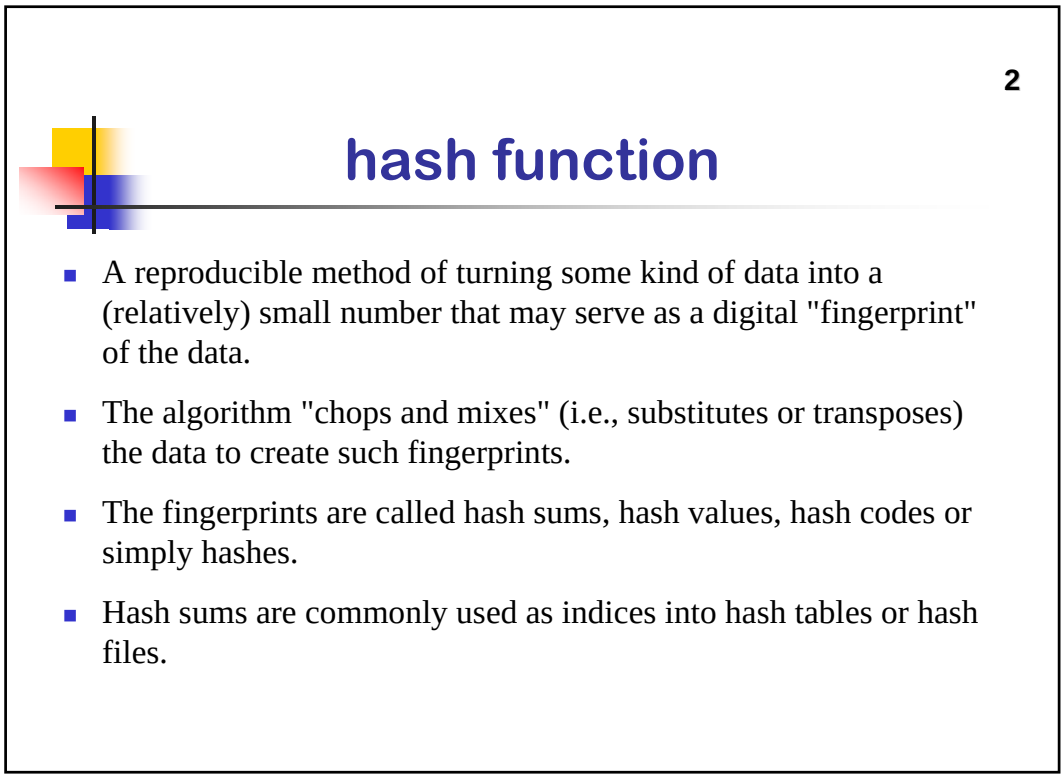
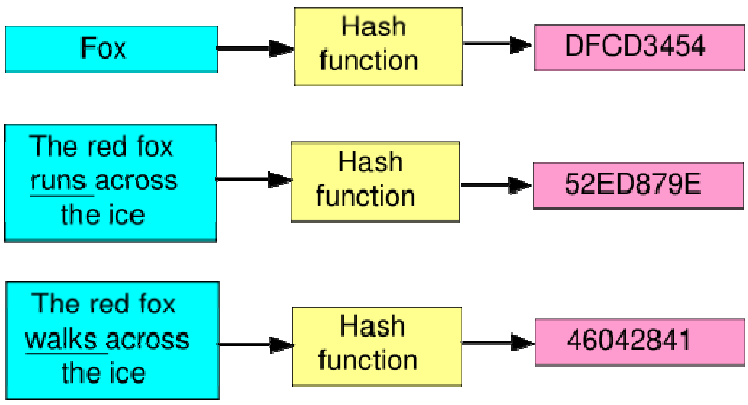

Hash Function
Dr.Talal Alkharobi


2
hash function
A reproducible method of turning some kind of data into a (relatively) small number that may serve as a digital "fingerprint" of the data.
The algorithm "chops and mixes" (i.e., substitutes or transposes) the data to create such fingerprints.
The fingerprints are called hash sums, hash values, hash codes or simply hashes.
Hash sums are commonly used as indices into hash tables or hash files.

3

Hash function

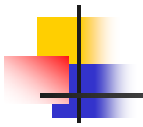


4

Hash functions

- Designed to be fast and to yield few hash collisions in expected input domains.
- In hash tables and data processing, collisions inhibit the distinguishing of data, making records more costly to find.
- A fundamental property of all hash functions is that if two hashes (according to the same function) are different, then the two inputs are different in some way.
- This property is a consequence of hash functions being deterministic.

5



Hash functions

- A function is not injective, i.e. the equality of two hash values ideally strongly suggests, but does not guarantee, the equality of the two inputs.
- If a hash value is calculated for a piece of data, and then one bit of that data is changed, a hash function with a strong mixing property usually produces a completely different hash value.
- Typical hash functions have an infinite domain, such as byte strings of arbitrary length, and a finite range, such as bit sequences of some fixed length.
- In certain cases, hash functions can be designed with one-to-one mapping between identically sized domain and range.

6

Hash functions

- Hash functions that are one-to-one are also called permutations.
- Reversibility is achieved by using a series of reversible "mixing" operations on the function input.

7

Applications

- Cryptography
- Hash tables
- Error correction
- Identification
- Verification

8

Cryptography hash function

- A hash function with certain additional security properties to make it suitable for use as a primitive in various information security applications, such as authentication and message integrity.
- A hash function takes a long string (or 'message') of any length as input and produces a fixed length string as output, sometimes termed a message digest or a digital fingerprint.
- In various standards and applications, the two most-commonly used hash functions are MD5 and SHA-1.

9

cryptographic hash function

- Broadly speaking, it should behave as much as possible like a random function while still being deterministic and efficiently computable.
- A cryptographic hash function is considered insecure if either of the following is computationally feasible:
  - Finding x given h(x)
  - Finding x & y such that h(x) = h(y)
- An attacker who can do either of these might use them to substitute an unauthorized message for an authorized one.

10

cryptographic hash function

- Ideally, it should not even be feasible to find two messages whose digests are substantially similar; nor would one want an attacker to be able to learn anything useful about x given only h(x)

11

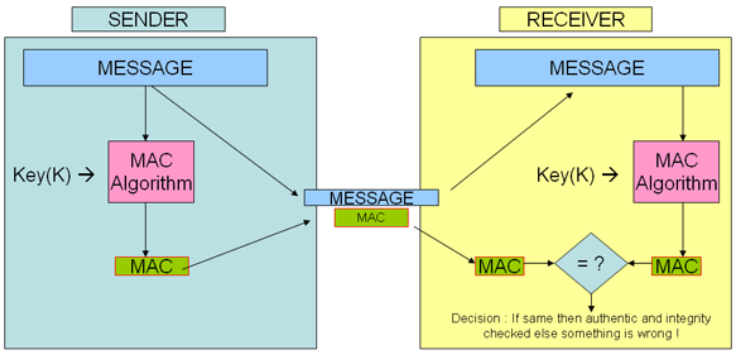
CRC and CryptoHash

- Checksums and cyclic redundancy checks (CRCs) are quite distinct from cryptographic hash functions, and are used for different applications.
- If used for security, they are vulnerable to attack;
  - A CRC was used for message integrity in the WEP encryption standard, but an attack was readily discovered which exploited the linearity of the checksum specified.

12

MAC

- A message authentication code or MAC takes a message and a secret key and generates a "MAC tag", such that it is difficult for an attacker to generate a valid pair (message, tag)



Message Authentication Code (MAC)

13

- A short piece of information used to authenticate a message.
- A MAC algorithm accepts as input a secret key and an arbitrary-length message to be authenticated, and outputs a MAC (sometimes known as a tag).
- The MAC value protects both a message's integrity as well as its authenticity, by allowing verifiers (who also possess the secret key) to detect any changes to the message content.
- A Message Integrity Code (MIC) is another name for a MAC.
- While MAC functions are similar to cryptographic hash functions, they possess different security requirements.

Message Authentication Code (MAC)

14

- To be considered secure, a MAC function must resist existential forgery under chosen-plaintext attacks.
- This means that even if an attacker has access to an oracle which possesses the secret key and generates MACs for messages of the attacker's choosing, he can "never" guess the MAC for any message that he has not yet asked the oracle about.
- Here "never" means, "not without doing an infeasible amount of computation".
- MACs differ from digital signatures, as MAC values are both generated and verified using the same secret key.

Message Authentication Code (MAC)

15

- This implies that the sender and receiver of a message must agree on keys before initiating communications, as is the case with symmetric encryption.
- For the same reason, MACs do not provide the property of non-repudiation offered by signatures: any user who can verify a MAC is also capable of generating MACs for other messages.
- In contrast, a digital signature is generated using the private key of a key pair, which is asymmetric encryption.
- Since this private key is only accessible to its holder, a digital signature proves that a document was signed by none other than that holder.

Message Authentication Code (MAC)

16

- MAC algorithms can be constructed from other cryptographic primitives, such as cryptographic hash functions (as in the case of HMAC) or from block cipher algorithms (OMAC, CBC-MAC and PMAC).

17

Cryptographic properties

- There is no formal definition which captures all desirable of a cryptographic hash function which are:
  - **Preimage resistant**: given x; it should be very hard to find any m such that x = h(m).
  - **Second preimage resistant**: given an input m, it should be very hard to find another input, n such that h(n) = h(m).
  - **Collision-resistant**: given h(m), it should be very hard to find a message n such that h(m) = h(n). Due to a possible birthday attack, the hash function output must be at least twice as large as what is required for preimage-resistance.

18

Cryptographic properties

- A hash function meeting these criteria may still have undesirable properties.
- For instance, most popular hash functions are vulnerable to length-extension attacks: given h(m) and len(m) but not m, by choosing a suitable n an attacker can calculate h (m || n), where || denotes concatenation.
- This property can be used to break naive authentication schemes based on hash functions.
- The HMAC construction works around these problems.

19

Hash algorithms

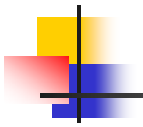
- Gost-Hash | HAS-160 | HAS-V | HAVAL | MDC-2 | MD2 | MD4 | MD5 | N-Hash | RadioGatún | RIPEMD | SHA family | Snefru | Tiger | VEST | WHIRLPOOL | crypt(3) DES

20

MD5

- Message-Digest algorithm 5 is a widely used cryptographic hash function with a 128-bit hash value.
- As an Internet standard (RFC 1321), MD5 has been employed in a wide variety of security applications, and is also commonly used to check the integrity of files.
- MD5 was designed by Ronald Rivest in 1991 to replace an earlier hash function, MD4.
- In 1996, a flaw was found with the design of MD5; while it was not a clearly fatal weakness, cryptographers began to recommend using other algorithms, such as SHA-1.
- In 2004, more serious flaws were discovered

21



MD5

- MD5 processes a variable-length message into a fixed-length output of 128 bits.
- The input message is broken up into chunks of 512-bit blocks; the message is padded so that its length is divisible by 512.
- The padding works as follows:
  - A 1is appended to the end of the message followed by as many zeros as are required to bring the length of the last block to 448 bits.
  - The remaining bits are filled up with a 64-bit integer representing the length of the original message.

22

MD5

- The main MD5 algorithm operates on a 128-bit state, divided into four 32-bit words, denoted A, B, C and D (initialized to certain fixed constants)
- The main algorithm then operates on each 512-bit message block in turn, each block modifying the state.
- The processing of a message block consists of four rounds;
- Each round is composed of 16 similar operations based on a non-linear function F, modular addition, and left rotation.
- There are four different functions F: F1, F2, F3, F4; different one is used for each round

23

MD5

- F1(X,Y,Z) = (*X* .and. *Y*) .or. (not.*X* .and. *Z*)
- F2(X,Y,Z) = (*X* .and. *Z*) .or. (not.*Z* .and. *Y*)
- F3(X,Y,Z) = *X* .xor. *Y* .xor. *Z*
- F4(X,Y,Z) = *Y* .xor. (*X* .or. not.*Z*)

24

MD5

- *Mi* denotes a 32-bit block of the message input,
- *Ki* denotes a 32-bit constant, different for each operation.
- <<<*s* denotes a left bit rotation by *s* places; *s* varies for each operation.
- The red squares denotes addition modulo 2^{32}.