CS 55: Security and Privacy

Hashing

Hash browns



Let's start with a game

Rules

You and I will both pick a number, if the sum is:

- Even I win
- Odd you win

Protocol

- Tell me your number!
- I will then tell you mine

Agenda

1. Hashing intro

- 2. Common hash functions
- 3. Hashing use cases
- 4. MACs and attacks

Back in CS10 we looked at hashing to find an index in table based on an object



Object



Table index

Goals:

- Compute quickly and consistently
- 2. Spreads keys over the table
- Small changes give different numbers

Step 1: convert object to integer

- Range –infinity to infinity
- Use hashCode()

Step 2: constrain to fall within hash table (hashCode %m)

Maps any object into table index from 0..m-1

Output is fixed length

Crypto hash functions add two properties: irreversibility and collision resistance



Hash function is a deterministic mathematical function

- 1. Irreversible Cannot find plain text in "reasonable" amount of time given only the hash digest output (one-way)
- Collision resistant different plaintext do not produce same hash digest
 Pseudorandom change any bit and get very different result

Let's visit the game with a new protocol

Rules

You and I will both pick a number, if the sum is:

- Even I win
- Odd you win

Protocol

- I hash my number and give you the result (commitment)
- You then tell me your number
- I then tell you my number
- You can confirm I didn't cheat by hashing my number and comparing with commitment
- Is this fair to both parties? Why or why not? What could be a problem with this protocol?



- 1. Hashing intro
- 2. Common hash functions
 - 3. Hashing use cases
 - 4. MACs and attacks

MD5 and SHA are popular choices, but SHA-2 is currently the most popular choice

MD5 (Message Digest 5)

- Popular choice of hash function
- Designed by Ron Rivest of RSA fame in 1991
- Collision resistance broken in 2004
- Can still use, but not if collision resistance is important

SHA (Secure Hash Algorithm)

- SHA-1
 - 160-bit hash function
 - Not recommended after 2005

SHA-2 the most common choice today

Don't use MD5 anymore for most purposes

- Collision found in 2017 (two different pdfs gave same hash)
- SHA-2
 - Two popular flavors: SHA-256 and SHA-512 (number = length of hash)
 - No known exploits... yet
- SHA-3
 - New as of 2015
 - Works differently from MD5 and SHA-2 (in case gets SHA2 broken)

MD5, SHA-1 and SHA-2 use the Merkle-Damgard construction

Merkle-Damgard



Message split into fixed sized blocks, processed one at a time

Fixed size current state initialized by IV

Each step uses the output of the prior step as input Digest is fixed length:

- MD5: 16 bytes (128 bits)
- SHA-256: 32 bytes (256 bits)
- SHA-512: 64 bytes (512 bits)

SHA-256 returns a 256-bit (32 byte) digest despite arbitrary length input

SHA-256

Break message in 512-bit blocks, pad last block

SHA256("abc")
Input: abc (string)
bytes: [97, 98, 99] #ascii values
Message block: 001100001011000101100011 #abc in binary

Pad message to make length a multiple of 512 bits:

- Add one bit
- Fill with zeros until last 64 bits
- Last 64 bits are length of message

Message block: 00110000101100010011000111000...11000 (length 24 bits = 3 bytes)

Message block is now 512 bits long

SHA256("abc")

= ba7816bf8f01cfea414140de5dae2223b00361a396177a9cb410ff61f20015ad

SHA-256 returns a 256-bit (32 byte) digest despite arbitrary length input

SHA-256

Simplified

Break message in 512-bit blocks, pad last block

- Calculate initial state of eight 32-bit variables based on square root of first 8 prime numbers
- Calculate constants K = cube root of first 64 primes **Compress**
- Split input in to 16 words of length 32 bits (512 bits) State (A—H) is 256 bits long
- Create 48 more words from first 16 words



- Changes made to state as message blocks fed in
- Math in compress is modulo 2³² so variables always stay at 32 bits
- Modulo looses information, making it hard to reverse

Round:

Ch(E,F,G) = $(E \land F) \oplus (-E \land G)$ Ma(A,B,C) = $(A \land B) \oplus (A \land C) \oplus (B \land C)$ $\Sigma_0(A) = (A \gg 2) \oplus (A \gg 13) \oplus (A \gg 22)$ $\Sigma_1(E) = (E \gg 6) \oplus (E \gg 11) \oplus (E \gg 25)$ \boxplus = addition modulo 32 Output A->B, B->C, ... A=Ch+Ma+ Σ_0 + Σ_1 %32 Repeat 63 times

SHA-256 returns a 256-bit (32 byte) digest despite arbitrary length input



Adapted from https://www.youtube.com/watch?v=DMtFhACPnTY, see https://youtu.be/f9EbD6iY9zI for detailed explanation of how SHA-2 works

SHA-256 returns a 256-bit (32 byte) digest despite arbitrary length input



Adapted from https://www.youtube.com/watch?v=DMtFhACPnTY, see https://youtu.be/f9EbD6iY9zI for detailed explanation of how SHA-2 works

Most Linux distributions come with utilities to compute hashes from the command line

#get md5 hash of plain.txt file #hashes to fixed length of 128 bits (16 bytes)

\$ md5sum plain.txt 9db228551f900dc17d6a8059bedc0880 plain.txt

#get SHA-256 hash of plain.txt file #hashes to fixed length of 256 bits (32 bytes)

\$ sha256sum plain.txt

45c946e93dc0509e4546f5489c1cd1083e7088976c7743c059ef70f49b031895 plain.txt

#get SHA-512 hash of plain.txt file #hashes to fixed length of 512 bits (64 bytes)

\$ sha512sum plain.txt

1969715ca7f29edcffaabcafa6aeaf4061a0d0b1bab8352e2d0da38c61437f5093ecfadfccc280e5 599a220da31c58fce7924841a036f82c889182b9551870fd plain.txt

#can also use openssl

\$ openssl dgst -sha512 plain.txt

SHA512(plain.txt)= 1969715ca7f29edcffaabcafa6aeaf4061a0d0b1bab8352e2d0da38c61437f5093ecfadfccc280e5 599a220da31c58fce7924841a036f82c889182b9551870fd 15

You can calculate hashes in Python using hashlib

compute_hash.py

try:

#set up hashlib for SHA512 m = hashlib.sha512()

#read data file line by line f = open(sys.argv[1],'r') for line in f: #update after each line m.update(line.encode("utf-8")) f.close() #print hash of file print(m.hexdigest()) except:

traceback.print_exc()

To confirm type "sha512sum <filename>" on command line

Note: same result if read line by line or read the whole file and then hash

You can also calculate hashes in C using OpenSSL

compute_hash.c

void main(int argc, char *argv[]) {
 SHA512_CTX ctx;
 u_int8_t results[SHA512_DIGEST_LENGTH];
 FILE * fp;
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 Run with: compute_hash plain.txt
 (Some code snipped for space on slide)
 (Some code snipped for space snipped for space on slide)
 (Some code snippe

SHA512_Init(&ctx); //set up for hashing

```
fp = fopen(argv[1], "r"); //open file for reading
if (fp == NULL) exit(EXIT_FAILURE);
```

```
while ((read = getline(&line, &len, fp)) != -1) { //read lines
    SHA512_Update(&ctx, line, strlen(line)); //update digest
}
```

SHA512_Final(results,&ctx); //finalize hashing

```
//print results
for(i=0;i< SHA512_DIGEST_LENGTH;i++) {
    printf("%02x",results[i]);
}
printf("\n");</pre>
```

Returns the same digest as Python

and the command line

Compile with: gcc compute_hash.c -o compute_hash -lcrypto

Some websites can give a hash such as SHA-256 or SHA-2512 on the fly

Website that calculates SHA hashes on the fly <u>https://www.movable-type.co.uk/scripts/sha512.html</u>



Movable Type Scripts

SHA-512 Cryptographic Hash Algorithm

A **cryptographic hash** (sometimes called 'digest') is a kind of 'signature' for a text or a data file. SHA-512 generates an almost-unique 512-bit (32-byte) signature for a text. See **below** for the source code.

Message message Hash f8daf57a3347cc4d6b9d575b31fe6077e2cb487f60a96233c08cb479dbf31538cc915ec6d48bd- baa96ddc1a16db4f4f96f37276cfcb3510b8246241770d5952c 0.430ms		[]
Hash f8daf57a3347cc4d6b9d575b31fe6077e2cb487f60a96233c08cb479dbf31538cc915ec6d48bd- baa96ddc1a16db4f4f96f37276cfcb3510b8246241770d5952c 0.430ms	lessage	message
	lash	f8daf57a3347cc4d6b9d575b31fe6077e2cb487f60a96233c08cb479dbf31538cc915ec6d48bd- baa96ddc1a16db4f4f96f37276cfcb3510b8246241770d5952c 0.430ms
Note SHA-512 hash of 'abc' should be: ddaf35a193617abacc417349ae20413112e6fa4e89a97ea20a9eeee64b55d39a219295	lote SHA-51	12 hash of 'abc' should be: ddaf35a193617abacc417349ae20413112e6fa4e89a97ea20a9eeee64b55d39a2192992a274fc1a6

This is a companion to the SHA-256 script (where there's more explanation). This is a reference implementation, as close as possible to the NIST specification, to help in understanding the algorithm (§ection numbers relate the code

Changing even one bit in the input message results in a different digest

#compute SHA-256 of message

\$ echo "Hello world!" | sha256sum 0ba904eae8773b70c75333db4de2f3ac45a8ad4ddba1b242f0b3cfc199391dd8

#change message and recompute hash

\$ echo "Hallo world!" | sha256sum bf1adae4567d9fb6b3bfb30cbf4dfdd2503e89a831cf3472c399b39fb9c73289

It is *extremely* unlikely hashes of two different inputs will collide

How unlikely you ask?

It is <u>extremely</u> unlikely that hashes of two different inputs will collide





- 1. Hashing intro
- 2. Common hash functions
- Hashing use cases
 - 4. MACs and attacks

Use case 1: Committing to a secret without revealing it

Committing to a secret

- Our game
- Hash secret and publish hash
- Disclosing hash does not disclose secret irreversible (one way)
- Once hash is published, cannot change document without being detected collision resistance

Example

- You've figured out a way to predict the stock market
- You don't want to tell everyone beforehand (or price goes up)
- Hash your predictions (with random nonce) and post them
- Later people can check your predictions were accurate

Use case 2: Integrity verification

Integrity verification (e.g., determine nothing has changed, or if it has!)

- Hash the original object
- To see if the object hash been changed, compare the original hash with a hash of the current version of the object

Example

Hash is equivalent to a duplicate copy because if anything changed in OS file, hashes won't match

- Want to know if OS kernel files have been changed by malware
- Could keep a spare copy of each OS file and check when going to run if it matches the spare copy – wasteful/impractical
- Instead, hash OS kernel files and store the 32-byte hash (not a second copy)
- Adversary might change kernel files, inserting malicious code
- Before running code, hash it and compare with stored version
- If no match, do not run it!

Many files on the Internet post a hash of the original Compare hash of what you downloaded with posted hash to confirm you got an unaltered copy (is this good enough?)

Password verification

- Do not store password in plaintext
- Store hashed version of password
- Use salt to defeat dictionary/rainbow tables
- To authenticate: enter password, hash it, compare with stored User name

Example

- In Linux passwords stored in /etc/shadow
- seed:\$6\$wDRrWCQz\$IsBXp9.9wz9SGrF.nbihpoN5w.zQx02sht4c TY8qI7YKh00wN/sfYvDeCAcEo2QYzCfpZoaEVJ8sbCT7hkxXY/:17 372:0:99999:7:::

Password verification

- Do not store password in plaintext
- Store hashed version of password
- Use salt to defeat dictionary/rainbow tables
- To authenticate: enter password, hash it, compare with stored User name

Example

6 means SHA512

- In Linux passwords stored in /etc/shadow
- seed:\$6\$wDRrWCQz\$IsBXp9.9wz9SGrF.nbihpoN5w.zQx02sht4c TY8qI7YKh00wN/sfYvDeCAcEo2QYzCfpZoaEVJ8sbCT7hkxXY/:17 372:0:99999:7:::

Password verification

- Do not store password in plaintext
- Store hashed version of password
- Use salt to defeat dictionary/rainbow tables
- To authenticate: enter password, hash it, compare with stored User name

Example

6 means SHA512

- In Linux passwords stored in /etc/shadow
- seed:\$6\$wDRrWCQz\$IsBXp9.9wz9SGrF.nbihpoN5w.zQx02sht4c TY8qI7YKh00wN/sfYvDeCAcEo2QYzCfpZoaEVJ8sbCT7hkxXY/:17 372:0:99999:7:::

Password verification

- Do not store password in plaintext
- Store hashed version of password
- Use salt to defeat dictionary/rainbow tables
- To authenticate: enter password, hash it, compare with stored User name

Example

🖌 6 means SHA

Password hash

- In Linux passwords stored in /etc/shadow
- seed:\$6\$wDRrWCQz\$IsBXp9.9wz9SGrF.nbihpoN5w.zQx02sht4c TY8qI7YKh00wN/sfYvDeCAcEo2QYzCfpZoaEVJ8sbCT7hkxXY/:17 372:0:99999:7:::

Password verification

- Do not store password in plaintext
- Store hashed version of password
- Use salt to defeat dictionary/rainbow tables
- To authenticate: enter password, hash it, compare with stored User name

Example

6 means SHA512 Salt

- Password hash
- In Linux passwords stored in /etc/shadow
- seed:\$6\$wDRrWCQz\$IsBXp9.9wz9SGrF.nbihpoN5w.zQx02sht4c TY8qI7YKh00wN/sfYvDeCAcEo2QYzCfpZoaEVJ8sbCT7hkxXY/:17 372:0:99999:7:::

\$ sudo cat /etc/shadow #list all passwords on system (requires sudo)	Does 5,000	
	rounds of	
seed:\$6\$wDRrWCQz\$IsBXp9.9wz9SGrF.nbihpoN5w.zQx02sht4cTY8qI7YKh00wN/	SHA512 to	
	make password	
<pre>\$ python3 make_linux_password_hash.py dees '\$6\$wDRrWCQz'</pre>	slow to	28
\$6\$wDRrWCQz\$IsBXp9.9wz9SGrF.nbihpoN5w.zQx02sht4cTY8qI7YKh00wN/	compute	20

Use case 4: Trusted timestamping

RFC 3161

Trusted timestamping

- Want to be able to prove a document existed at a point in time (called Long Term Validation – LTV)
- Also want non-repudiation (cannot later deny something, repudiate means to deny)

Use case 4: Trusted timestamping

RFC 3161

Trusted timestamping

- Want to be able to prove a document existed at a point in time (called Long Term Validation – LTV)
- Also want non-repudiation (cannot later deny something, repudiate means to deny)



Discussion

Scenario:

- You are pitching a start up idea to a sketchy venture capitalist
- You'd like to be able to prove you have produced your business plan *before* you meet with them
- What can you do?



- 1. Hashing intro
- 2. Common hash functions
- 3. Hashing use cases
- 4. MACs and attacks

We need to guard against a message that has been changed or forged

Alice sends message to Bob



How does Bob know if the message has been altered?

Message Authentication Code (MAC)

Not secure!

Alice

This looks promising (assuming Bob knows key and adversary does not)

Message: Launch missile at Target A MAC: Hash(key||message)

Changing the message fed into a hash

function by even one bit changes output



Bob

Assume Alice and Bob share a secret key

Alice could hash(key || message) and send hash with message (called a Message Authentication Code, MAC, or a tag) How does Bob know if the message has been altered?

Bob knows Key and can hash key||message and compare with MAC sent by Alice

Message Authentication Code (MAC)

Not secure!



Assume Alice and Bob share a secret key

How does Bob know if the message has been altered?

Alice could *hash(key || message)* and send hash with message (called a Message Authentication Code, MAC, or a tag)

echo -n "theKey:Launch a missile at Target A" | sha256sum 5a97c8de6b858a3bb145b62661bb511eb7c77aeb1ef0cf86d585d8b81ecef2e7

Message Authentication Code (MAC)

Not secure!



Assume Alice and Bob share a secret key

Alice could *hash(key || message)* and send hash with message (called a Message Authentication Code, MAC, or a tag) How does Bob know if the message has been altered?

Bob knows Key and can hash key||message and compare with MAC

36

echo -n "theKey:Launch a missile at Target A" | sha256sum 5a97c8de6b858a3bb145b62661bb511eb7c77aeb1ef0cf86d585d8b81ecef2e7

Message Authentication Code (MAC)

Not secure!



echo -n "theKey:Launch a missile at Target A" | sha256sum 5a97c8de6b858a3bb145b62661bb511eb7c77aeb1ef0cf86d585d8b81ecef2e7 37

Message Authentication Code (MAC)

Not secure!



Hash length extension attacks exploit Merkle-Damgard construction

Hash length extension attack



Because a block depends on the prior block

- If adversary intercepts message with MAC and can guess key length
- Adversary can add more text on to the end and still get a valid MAC, even though they do not know the key

Hash length extension attacks exploit Merkle-Damgard construction

Hash length extension attack



- Pad with 1 followed by zeros,
- Pad ends with size of key:message in hex
 - Len("theKey:Launch missile at Target A") = 35
 - End with 35 bytes * 8 = 280 (decimal) = 0x118 (hex)

This is the value that is actually hashed

Message Authentication Code (MAC)

Not secure!



Adversary resets SHA256 context to where it left off, then adds new message Legitimate message: "Launch a missile at Target A"

sha256_length_extension.c



for (i =0; i<64; i++) SHA256 Update(&c, "*", 1);



Reset context to where hashing left off using MAC

> Add new message (this is the message length extension)

// Append the additional message SHA256_Update(&c, " and at Hanover", 15); SHA256 Final(buffer, &c); **Print new MAC** Note: padding not printable characters! for (i = 0; i < 32; i++) { Send Bob printf("%02x", buffer[i]); 'Launch a missile at Target A <padding> and at Hanover" printf("\n"); 42 Adapted from Du, Wenliang. Computer & Internet Security: A Hands-on Approach. 2019.

MAC checks out at the receiver, even though adversary does not know secret

Message: "Launch a missile at Target A

sha256_padding.c

```
<padding> and at Hanover"
int main(int argc, const char *argv[]) {
                                        New MAC: bebb350f2613abff0520fa9754cc4
SHA256 CTX c;
                                       4cb58d36e3ec17fbd0092a2e9ecd56a0792
unsigned char buffer[SHA256 DIGEST LENGTH];
int i:
SHA256 Init(&c);
                                 Set up SHA256 context
SHA256 Update(&c,
                                                  Hash with original message,
 "theKey:Launch a missile at Target A"
                                                  padding, and extension
 " and at Hanover",
 64+17):
SHA256 Final(buffer, &c);
                                   MAC matches the one sent
printf("New MAC\n");
for (int i = 0; i < 32; i++) {
 printf("%02x", buffer[i]);
                              Two missiles launched!
printf("\n");
return 0:
```

```
Adapted from Du, Wenliang. Computer & Internet Security: A Hands-on Approach. 2019.
```

A better solution is to use a Keyed-Hash Message Authentication Code (HMAC)

- HMAC (K,m) = H((K \oplus opad) || H((K \oplus ipad) || m))
- Requires secret key and two hashes using:
- K = key
- opad = outer pad = 0x5c5c5c5c
- ipad = inner pad = 0x36363636



A better solution is to use a Keyed-Hash Message Authentication Code (HMAC)

Key

HMAC (K,m) = H((K \oplus opad) || H((K \oplus ipad) || m))

Requires secret key and two hashes using:

- K = key
- opad = outer pad = 0x5c5c5c5c
- ipad = inner pad = 0x36363636

Inner hash uses irreversibility property where adversary without key cannot find message

Output is a fingerprint on the message



Adapted from Du, Wenliang. Computer & Internet Security: A Hands-on Approach. 2019.

and https://crypto.stackexchange.com/questions/12680/how-does-the-secret-key-in-an-hmac-prevent-modification-of-the-hmac