COMP8760 Lecture

Hash Functions

Sanjay Bhattacherjee

University of Kent



Outline

Hash Functions: Preliminaries

"Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding The Merkle–Damgård construction SHA-256





Study Material for Hash Functions

Book 1 Cryptography Made Simple
Author Nigel P. Smart.
Link to eBook

Section 1.4.2 Birthday Paradox Section 1.5 Big Numbers

Chapter 14 Hash Functions
Section 14.1 Collision Resistance
Section 14.2 Padding
Section 14.3 The Merkle–Damgård construction
Section 14.4.3 SHA-256

Book Sections

Part 1 Preliminaries

- Birthday Paradox (Section 1.4.2)
- Big Numbers (Section 1.5)

Part 2 Cryptographic Hash Function

- The three security properties (Section 14.1)
- Padding (Section 14.2)
- The Merkle–Damgård construction (Section 14.3)
- SHA-256 (Section 14.4.3)



Outline

Hash Functions: Preliminaries

Birthday Bound "Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding The Merkle–Damgård construction SHA-256





What is the probability that none of them have the same birthday?

What is the probability that none of them have the same birthday?



What is the probability that none of them have the same birthday?



What is the probability that none of them have the same birthday?

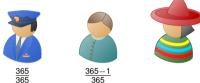




What is the probability that none of them have the same birthday?

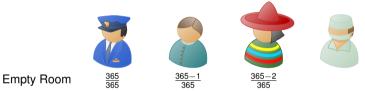




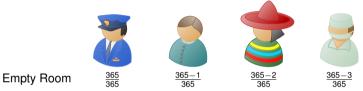




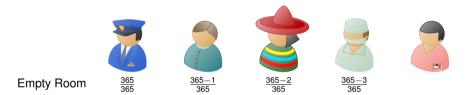




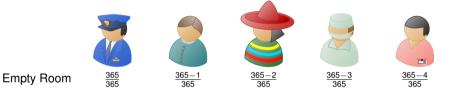




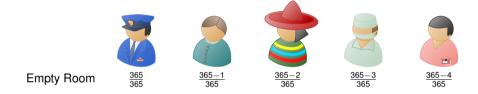












$$\Pr[\text{None have same birthday}] = \frac{m \cdot (m-1) \cdot (m-2) \cdots (m-k+1)}{m^k} = \frac{m_{P_k}}{m^k}$$

$$\Pr[\text{At least one pair collides}] = 1 - \frac{m \cdot (m-1) \cdot (m-2) \cdots (m-k+1)}{m^k} = 1 - \frac{m_{P_k}}{m^k}$$



Birthday Paradox: Probability of collision ≈ 1 for $k \ll 365$

What is the probability that none of them have the same birthday?



▶ With 23 people in a room, the probability of a collision is

$$1 - \frac{^{365}P_{23}}{365^{23}} = 0.507.$$

▶ With 30 people in a room, the probability of a collision is

$$1 - \frac{^{365}P_{30}}{365^{30}} = 0.706.$$

▶ With 100 people in a room, the probability of a collision is

$$1 - \frac{^{365}P_{100}}{365^{100}} > 0.999.$$





Birthday Bound

The probability of collision in choosing n elements from m is at most

$$\frac{n^2}{2 \cdot m}$$











Outline

Hash Functions: Preliminaries

Birthday Bound

"Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding

The Merkle-Damgård construction

SHA-256



Computational Difficulty

How big can a number be before it is impossible for someone to perform that many (classical) computer operations.

- $ightharpoonup 2^{10} \approx 10^3$.
- ▶ 3GHz processor can execute 3,000,000,000 instructions per second.
- ► 3GHz octa-core processor can execute 24 × 10⁹ instructions per second.
- Let $\mathcal C$ be an advanced classical computer that performs $10^{12}\approx 2^{40}$ instructions per second.
 - For 2⁶⁴ operations:
 \$\mathcal{C}\$ would take \$\frac{2^{64}}{2^{40}}\$ seconds or 194 days;
 or one day for 194 \$\mathcal{C}\$s.
 For 2⁸⁰ operations (> 10²³):
 - For 200 operations (> 10^{23}): C would take $\frac{2^{90}}{2^{40}}$ seconds or 34,900 years;
 or 2 years for 15,000 Cs.

 Still imaginable!
 - For 2^{128} operations (> 10^{37}): C would take $\frac{2^{128}}{2^{40}}$ seconds or $9 \cdot 10^{18}$ years. Unimaginable!





Outline

Hash Functions: Preliminaries

"Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding The Merkle–Damgård construction SHA-256





Cryptographic Hash Function

Hash Function

A function that takes arbitrary length bit strings as input and produces a fixed-length bit string as output;

The output is often called digest, hashcode or hash value.

Cryptographic Hash Function

Hash Function

A function that takes arbitrary length bit strings as input and produces a fixed-length bit string as output;

The output is often called digest, hashcode or hash value.

Cryptographic Hash Function

... is a hash function that is one-way.

In other words, given $H:D\to C$, and $y\in C$, it is computationally infeasible to find any value $x\in D$ such that

$$y = H(x)$$
.

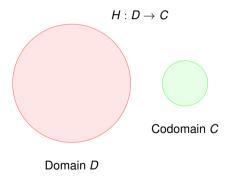




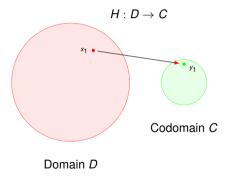
► Long input, fixed-length output (hash / digest)

- ► Long input, fixed-length output (hash / digest)
- ► Easy to compute

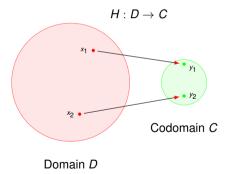
- ► Long input, fixed-length output (hash / digest)
- Easy to compute
- ► Hard to invert



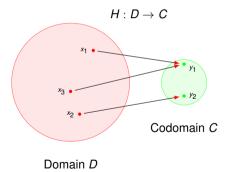
- ► Long input, fixed-length output (hash / digest)
- Easy to compute
- ► Hard to invert



- ► Long input, fixed-length output (hash / digest)
- Easy to compute
- ► Hard to invert



- ► Long input, fixed-length output (hash / digest)
- Easy to compute
- ► Hard to invert





Applications



- Integrity checks: if the hashes do not match, the message must have been altered
- Authentication: through cryptographic signatures
- ► Key derivation functions: to generate keys from various inputs
- Proof-of-work: blockchains, mitigating spamming
- Password: stored in place of the plaintext password
- Indexing for search: in databases, other file systems
- Perceptual hashing: proximity search for text/image/video
- et cetera ...





Outline

Hash Functions: Preliminaries

"Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

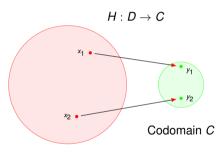
Padding
The Merkle–Damgård construction
SHA-256





1. Preimage resistance

For a hash function H which produces outputs of t bits, finding preimages should require $O(2^t)$ time.



Domain D

Preimage Resistance \equiv One-Way Function

Challenger

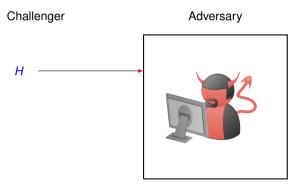
Adversary



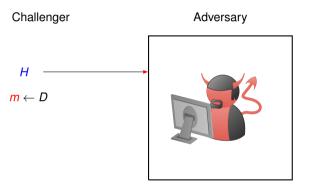
Advantage of Adversary = Pr[Adversary wins the game]



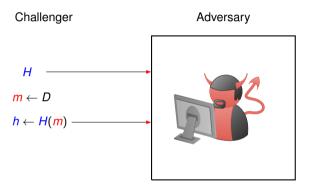
Preimage Resistance \equiv One-Way Function



Preimage Resistance \equiv One-Way Function

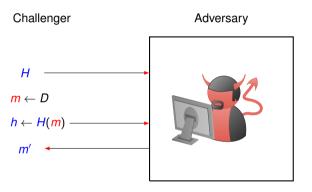


Preimage Resistance ≡ One-Way Function



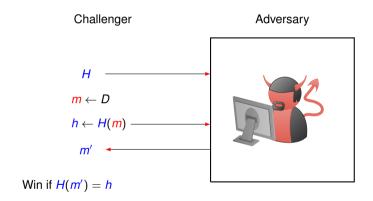


Preimage Resistance \equiv One-Way Function





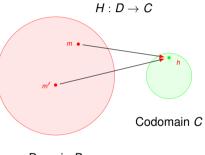
Preimage Resistance ≡ One-Way Function





2. Second-preimage resistance

Given m, it should be hard to find an $m' \neq m$ with H(m') = H(m).

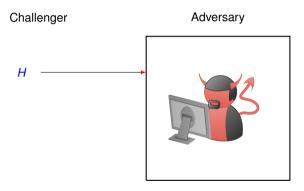


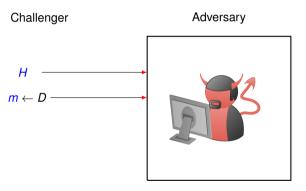
Challenger

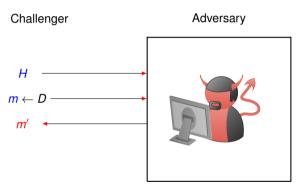
Adversary

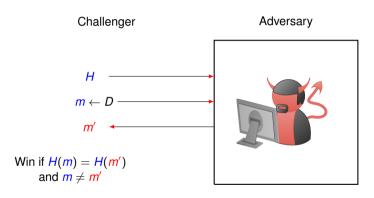






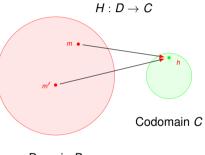








It should be hard to find a pair $m' \neq m$ with H(m') = H(m).



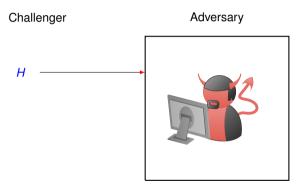
Domain D

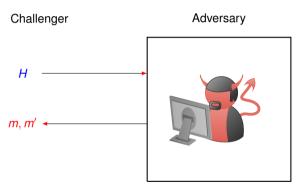
Challenger

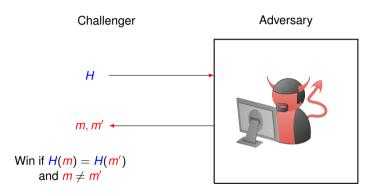
Adversary













$$H: D \rightarrow C$$

 $H: D \rightarrow C$

A function H is said to be collision resistant (by human ignorance) or HI-CR secure

$$H: D \rightarrow C$$

A function *H* is said to be collision resistant (by human ignorance) or HI-CR secure if it is believed to be infeasible to find a collision.

$$H: D \rightarrow C$$

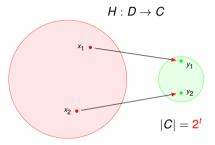
A function *H* is said to be collision resistant (by human ignorance) or HI-CR secure if it is believed to be infeasible to find a collision.

In other words, it is infeasible to find two elements $m, m' \in D$ such that

$$H(m) = H(m').$$

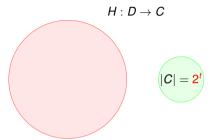


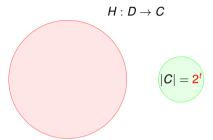
How many queries to find a preimage?



For a "well-designed hash function", the preimage of an element would be one of 2^t elements of D. So, the number of queries would be approximately

$$q_{\text{finding preimage}} = 2^t$$
.

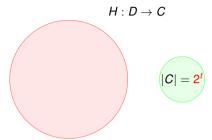




By the birthday bound, after q queries, the probability of collision is at most

$$\frac{q^2}{2 \cdot 2^q}$$



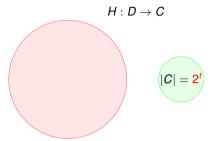


By the birthday bound, after q queries, the probability of collision is at most

$$\frac{q^2}{2 \cdot 2^t}$$

So, a collision will almost certainly be found after

$$q_{\text{finding collision}} = \sqrt{2^{t+1}}$$
 queries.



By the birthday bound, after q queries, the probability of collision is at most

$$\frac{q^2}{2 \cdot 2^t}$$

So, a collision will almost certainly be found after

$$q_{\text{finding collision}} = \sqrt{2^{t+1}}$$
 queries.

So, if we require 128-bit security of a hash function, we would want the output to be of size at least 256 bits.

Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage.

Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage. We construct an adversary $\mathcal{A}_{collision}$ for finding a collision.



Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage. We construct an adversary $\mathcal{A}_{collision}$ for finding a collision.

Let $\mathcal{O}_{preimage}$ be an oracle that provides a preimage for any element of C

Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage. We construct an adversary $\mathcal{A}_{collision}$ for finding a collision.

- Let $\mathcal{O}_{preimage}$ be an oracle that provides a preimage for any element of C
- ► Choose a random $x \in D$ and find h = H(x)



Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage. We construct an adversary $\mathcal{A}_{collision}$ for finding a collision.

- Let $\mathcal{O}_{preimage}$ be an oracle that provides a preimage for any element of C
- ▶ Choose a random $x \in D$ and find h = H(x)
- ▶ Find $x' \leftarrow \mathcal{O}_{preimage}(h)$

Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage. We construct an adversary $A_{collision}$ for finding a collision.

- Let $\mathcal{O}_{preimage}$ be an oracle that provides a preimage for any element of C
- ▶ Choose a random $x \in D$ and find h = H(x)
- ▶ Find $x' \leftarrow \mathcal{O}_{preimage}(h)$

Since *D* is a huge set with many preimages of *h*, hence it is highly likely that $x \neq x'$.

Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage. We construct an adversary $A_{collision}$ for finding a collision.

- Let $\mathcal{O}_{preimage}$ be an oracle that provides a preimage for any element of C
- ▶ Choose a random $x \in D$ and find h = H(x)
- ▶ Find $x' \leftarrow \mathcal{O}_{preimage}(h)$

Since D is a huge set with many preimages of h, hence it is highly likely that $x \neq x'$. Thus, $\mathcal{A}_{collision}$ has found a collision using $\mathcal{O}_{preimage}$!

If there is an efficient $\mathcal{O}_{\textit{preimage}}$, then there is an efficient $\mathcal{A}_{\textit{collision}}$. By contrapositivity, since there is no efficient $\mathcal{A}_{\textit{collision}}$, hence there is no efficient $\mathcal{O}_{\textit{preimage}}$.

Lemma 14.2

If a function H is preimage resistant for every element of the range of H, it will be collision resistant and second-preimage resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a preimage. We construct an adversary $A_{collision}$ for finding a collision.

- Let $\mathcal{O}_{preimage}$ be an oracle that provides a preimage for any element of C
- ▶ Choose a random $x \in D$ and find h = H(x)
- ▶ Find $x' \leftarrow \mathcal{O}_{preimage}(h)$

Since D is a huge set with many preimages of h, hence it is highly likely that $x \neq x'$. Thus, $\mathcal{A}_{collision}$ has found a collision using $\mathcal{O}_{preimage}$!

If there is an efficient $\mathcal{O}_{\textit{preimage}}$, then there is an efficient $\mathcal{A}_{\textit{collision}}$. By contrapositivity, since there is no efficient $\mathcal{A}_{\textit{collision}}$, hence there is no efficient $\mathcal{O}_{\textit{preimage}}$.

Similarly for second-preimage resistance.

Hence, Preimage Resistance is a weaker assumption than Collision Resistance or Second-Preimage Resistance.





Lemma 14.3

If a function H is second-preimage resistant, then it is also collision resistant.



Lemma 14.3

If a function H is second-preimage resistant, then it is also collision resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a second-preimage.

Lemma 14.3

If a function H is second-preimage resistant, then it is also collision resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a second-preimage. We construct an adversary $\mathcal{A}_{\textit{collision}}$ for finding a collision.

Lemma 14.3

If a function H is second-preimage resistant, then it is also collision resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a second-preimage. We construct an adversary $A_{collision}$ for finding a collision.

▶ Let $O_{second-preimage}$ be an oracle that given an input $x \in D$, provides a second-preimage $x' \in D$.

Lemma 14.3

If a function H is second-preimage resistant, then it is also collision resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a second-preimage. We construct an adversary $A_{collision}$ for finding a collision.

- ▶ Let $O_{second-preimage}$ be an oracle that given an input $x \in D$, provides a second-preimage $x' \in D$.
- ▶ Choose a random $x \in D$

Lemma 14.3

If a function H is second-preimage resistant, then it is also collision resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a second-preimage. We construct an adversary $A_{collision}$ for finding a collision.

- ▶ Let $O_{second-preimage}$ be an oracle that given an input $x \in D$, provides a second-preimage $x' \in D$.
- ▶ Choose a random $x \in D$
- $ightharpoonup x' \leftarrow \mathcal{O}_{second-preimage}(x)$

Thus, $A_{collision}$ has found a collision using $O_{second-preimage}$!

Lemma 14.3

If a function H is second-preimage resistant, then it is also collision resistant.

Proof

We prove this by reducing the problem of finding a collision to that of finding a second-preimage. We construct an adversary $A_{collision}$ for finding a collision.

- ▶ Let $O_{second-preimage}$ be an oracle that given an input $x \in D$, provides a second-preimage $x' \in D$.
- ightharpoonup Choose a random $x \in D$
- $\triangleright x' \leftarrow \mathcal{O}_{second-preimage}(x)$

Thus, $A_{collision}$ has found a collision using $O_{second-preimage}$!

If there is an efficient $\mathcal{O}_{second-preimage}$, then there is an efficient $\mathcal{A}_{collision}$. By contrapositivity, since there is no efficient $\mathcal{A}_{collision}$, hence there is no efficient $\mathcal{O}_{second-preimage}$.

Hence, Second-Preimage Resistance is a weaker assumption than Collision Resistance.



Outline

Hash Functions: Preliminaries

"Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding The Merkle–Damgård constructior SHA-256





Outline

Hash Functions: Preliminaries

"Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding

The Merkle–Damgård construction SHA-256





What if the entire message does not fit in one plaintext?

What if the entire message does not fit in one plaintext?

▶ Split the message into smaller plaintext blocks and encrypt each block.

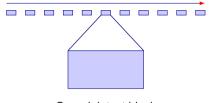


One plaintext block



What if the entire message does not fit in one plaintext?

Split the message into smaller plaintext blocks and encrypt each block.



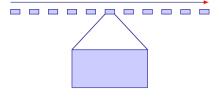
One plaintext block

What if the number of bits in the message is not a multiple of the number of bits in one plaintext block?



What if the entire message does not fit in one plaintext?

Split the message into smaller plaintext blocks and encrypt each block.



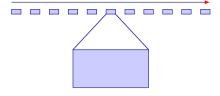
One plaintext block

- What if the number of bits in the message is not a multiple of the number of bits in one plaintext block?
 - The block size **b** = number of bits in one block



What if the entire message does not fit in one plaintext?

Split the message into smaller plaintext blocks and encrypt each block.



One plaintext block

- What if the number of bits in the message is not a multiple of the number of bits in one plaintext block?
 - The block size **b** = number of bits in one block
 - Add some redundant bits to the message so that.

$$|\mathbf{m}| = \mathbf{k} \cdot \mathbf{b}$$
.

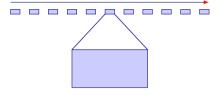






What if the entire message does not fit in one plaintext?

▶ Split the message into smaller plaintext blocks and encrypt each block.



One plaintext block

- What if the number of bits in the message is not a multiple of the number of bits in one plaintext block?
 - The block size b =number of bits in one block
 - Add some redundant bits to the message so that.

$$|\mathbf{m}| = \mathbf{k} \cdot \mathbf{b}.$$

The manner of adding bits should allow the decryption algorithm to separate the portion of the original plaintext from the padded part of the plaintext block.



m
$$pad_0(|m|, b)$$
Also written as: $\underline{m||pad_0(|m|, b)}$
Concatenation

Let
$$v \leftarrow b - \underbrace{|m| \pmod{b}}$$
. Then, the padded message is

$$m||\underbrace{pad_0(|m|,b)}_{padding \ bits} = m||\underbrace{0...0}_{v \ bits}.$$



m
$$pad_1(|m|, b)$$
Also written as: $\underline{m||pad_1(|m|, b)}$
Concatenation

Method 1:

Let
$$v \leftarrow b - \underbrace{(|m|+1) \pmod{b}}_{\text{message bits} + 1 \text{ in last block}}$$
 Then, the padded message is

$$|m||\underbrace{pad_1(|m|,b)}_{padding bits} = |m||1\underbrace{0...0}_{v \text{ bits}}.$$



$$m$$
 $pad_2(|m|, b)$

Also written as:
$$\underline{\underline{m||pad_2(|m|,b)}}$$
Concatenation

Method 2:

Let
$$v \leftarrow b - \underbrace{(|m| + 64 + 1) \pmod{b}}$$
. Then, the padded message is message bits +65 in last block

$$m||\underbrace{\textit{pad}_2(|m|,b)}_{\text{padding bits}} = m||1\underbrace{0\ldots0}_{\text{v bits}}||\underbrace{|m|}_{\text{64 bits}}.$$



m
$$pad_3(|m|, b)$$
Also written as: $\underline{m||pad_3(|m|, b)}$
Concatenation

Method 3: Let $v \leftarrow b - (|m| + 64) \pmod{b}$. Then, the padded message is message bits +64 in last block

$$m||\underbrace{pad_3(|m|,b)}_{\text{padding bits}} = m||\underbrace{0...0}_{v \text{ bits}}||\underbrace{|m|}_{64 \text{ bits}}$$





m
$$pad_4(|m|, b)$$
Also written as: $m||pad_4(|m|, b)$

Let
$$v \leftarrow b - \underbrace{(|m|+2) \pmod{b}}$$
. Then, the padded message is message bits +2 in last block

$$m||\underbrace{\mathit{pad}_4(|m|,b)}_{\mathsf{padding \, bits}} = m||1\underbrace{0\ldots 0}_{\mathit{v \, bits}} 1.$$

Outline

Hash Functions: Preliminaries Birthday Bound "Big" Numbers

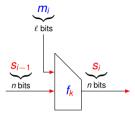
Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding
The Merkle–Damgård construction
SHA-256



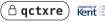


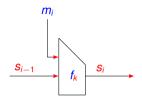
The Building Block: A Compression Function

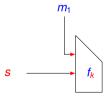
► A compression function is a hash function taking inputs of a fixed length:

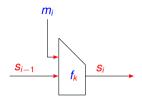
$$f_k: \{0,1\}^{\ell+n} \to \{0,1\}^n.$$

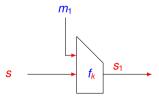
- ▶ Map $(\ell + n)$ -bit inputs to n-bit outputs.
- ightharpoonup When applied iteratively, the variable s_i works as an internal state.

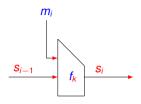


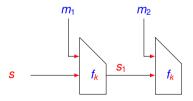


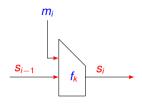


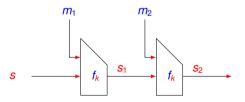


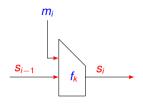


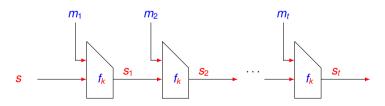


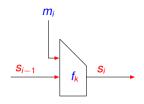




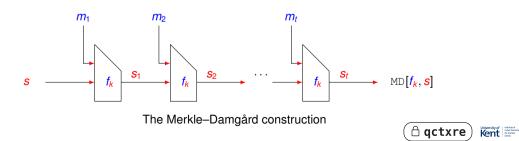








The Building Block: A Compression Function



The Algorithm

Merkle-Damgård construction

The Algorithm

Merkle-Damgård construction

Step 1: Padding: $m||pad_i(|m|, \ell)$

The Algorithm

Merkle-Damgård construction

Step 1: Padding: $m||pad_i(|m|, \ell)$

Step 2: Split $\underline{m||pad_i(|m|, \ell)}$ into $\underline{m_1||m_2||\cdots||m_t}$, each of length ℓ bits.

The Algorithm

Merkle-Damgård construction

Step 1: Padding: $m||pad_i(|m|, \ell)$

Step 2: Split $\underline{m||pad_i(|m|, \ell)}_{padded \ message}$ into $\underline{m_1||m_2||\cdots||m_t}_{t \ blocks}$, each of length ℓ bits.

Step 3: $s_0 \leftarrow s$

The Algorithm

Merkle-Damgård construction

```
Step 1: Padding: m||pad_i(|m|, \ell)
```

Step 2: Split
$$\underline{m||pad_i(|m|, \ell)}$$
 into $\underline{m_1||m_2||\cdots||m_t}$, each of length ℓ bits.

Step 3:
$$s_0 \leftarrow s$$

Step 4: For
$$i = 1$$
 to t do:
 $f(s_{i-1}||m_i) \rightarrow s_i$

The Algorithm

Merkle-Damgård construction

```
Step 1: Padding: m||pad_i(|m|, \ell)
```

Step 2: Split
$$\underline{m||pad_i(|m|, \ell)}$$
 into $\underline{m_1||m_2||\cdots||m_t}$, each of length ℓ bits.

Step 3: $s_0 \leftarrow s$

Step 4: For
$$i = 1$$
 to t do:
 $f(s_{i-1}||m_i) \rightarrow s_i$

Step 5: return s_t

Padding Scheme determines Collision Resistance

Let $n = \ell = 4$. Consider the messages

$$m_1 = 0b0$$
 and $m_2 = 0b00$.



Padding Scheme determines Collision Resistance

Let $n = \ell = 4$. Consider the messages

$$m_1 = 0b0$$
 and $m_2 = 0b00$.

When padded using Method 0, we get the same message:

0*b*000000.



Padding Scheme determines Collision Resistance

Let $n = \ell = 4$. Consider the messages

$$m_1 = 0b0$$
 and $m_2 = 0b00$.

When padded using Method 0, we get the same message:

0*b*000000.

So, the output of f(0b000000) will of course be equal in both cases, resulting in a collision!



The Merkle-Damgård construction

Padding Scheme determines Collision Resistance

Let $n = \ell = 4$. Consider the messages

$$m_1 = 0b0$$
 and $m_2 = 0b00$.

When padded using Method 0, we get the same message:

0*b*000000.

So, the output of f(0b000000) will of course be equal in both cases, resulting in a collision!

Theorem 14.4

Let

$$H_{k,s}(m) = MD[f_{k,s}](m)$$

denote the keyed hash function constructed using the Merkle–Damgård method from the keyed compression function $f_k(x)$ and Padding Method 2.

If $f_k(x)$ is collision resistant, then $H_{k,s}$ is collision resistant as well.





Commonly used Hash functions

► MD-5, RIPEMD-160, SHA-1 and SHA-2

- ► MD-5, RIPEMD-160, SHA-1 and SHA-2
 - ► All are based on the Merkle-Damgård construction
 - ► All use a fixed (i.e. unkeyed) compression function f



- ► MD-5, RIPEMD-160, SHA-1 and SHA-2
 - ► All are based on the Merkle–Damgård construction
 - ► All use a fixed (i.e. unkeyed) compression function *f*
- ► MD-5: outputs 128-bit hashes



- MD-5, RIPEMD-160, SHA-1 and SHA-2
 - ► All are based on the Merkle–Damgård construction
 - ► All use a fixed (i.e. unkeyed) compression function *f*
- ► MD-5: outputs 128-bit hashes
- ► RIPEMD-160 and SHA-1: output 160-bit hashes



- ► MD-5, RIPEMD-160, SHA-1 and SHA-2
 - ► All are based on the Merkle–Damgård construction
 - ► All use a fixed (i.e. unkeyed) compression function f
- ► MD-5: outputs 128-bit hashes
- ► RIPEMD-160 and SHA-1: output 160-bit hashes
- ► SHA-2: three algorithms

- ► MD-5, RIPEMD-160, SHA-1 and SHA-2
 - ► All are based on the Merkle–Damgård construction
 - ► All use a fixed (i.e. unkeyed) compression function *f*
- ► MD-5: outputs 128-bit hashes
- RIPEMD-160 and SHA-1: output 160-bit hashes
- SHA-2: three algorithms
 - SHA-256: outputs 256-bit hashes
 - ► SHA-384: outputs 384-bit hashes
 - ► SHA-512: outputs 512-bit hashes



Outline

Hash Functions: Preliminaries

"Big" Numbers

Cryptographic Hash Function

Hash Function: Security Properties

SHA-256

Padding
The Merkle—Damgård construction
SHA-256





▶ Input: message m of length $0 \le |m| < 2^{64}$ to be divided into t blocks

$$\underbrace{\frac{m_1}{512 \text{ bits}}, \underbrace{\frac{m_2}{512 \text{ bits}}}_{512 \text{ bits}}, \underbrace{\frac{m_t}{512 \text{ bits}}}_{512 \text{ bits}}$$

each of length $\ell = 512$ bits.

▶ Input: message m of length $0 \le |m| < 2^{64}$ to be divided into t blocks

$$\underbrace{\frac{m_1}{512 \text{ bits}}, \frac{m_2}{512 \text{ bits}}, \dots, \underbrace{m_t}_{512 \text{ bits}}}_{\leq 2^{64} \approx 16 \times 10^{24} \text{ bits}}$$

each of length $\ell = 512$ bits.

For each m_i , the $\ell = 512$ bits are further sub-divided into 16 words

$$\underbrace{X_1}_{32 \text{ bits}}, \underbrace{X_2}_{32 \text{ bits}}, \underbrace{X_2}_{32 \text{ bits}}, \dots, \underbrace{X_{16}}_{32 \text{ bits}},$$

each of size 32 bits.



▶ Input: message m of length $0 \le |m| < 2^{64}$ to be divided into t blocks

$$\underbrace{\frac{m_1}{512 \text{ bits}}, \frac{m_2}{512 \text{ bits}}, \dots, \frac{m_t}{512 \text{ bits}}}_{\leq 2^{64} \approx 16 \times 10^{24} \text{ bits}}$$

each of length $\ell = 512$ bits.

For each m_i , the $\ell = 512$ bits are further sub-divided into 16 words

$$\underbrace{X_1}_{32 \text{ bits}}, \underbrace{X_2}_{32 \text{ bits}}, \underbrace{X_2}_{32 \text{ bits}}, \dots, \underbrace{X_{16}}_{32 \text{ bits}}$$

each of size 32 bits.

▶ Padding: Method 2 (to make sure |m| is a multiple of $\ell = 512$)



▶ Input: message m of length $0 \le |m| < 2^{64}$ to be divided into t blocks

$$m_1$$
, m_2 , ..., m_t
 512 bits 512 bits 512 bits 512 bits $\leq 2^{64} \approx 16 \times 10^{24} \text{ bits}$

each of length $\ell = 512$ bits.

For each m_i , the $\ell=512$ bits are further sub-divided into 16 words

$$\underbrace{X_1}_{32 \text{ bits } 32 \text{ bits } 32 \text{ bits}}, \underbrace{X_2}_{32 \text{ bits } 32 \text{ bits}}, \underbrace{X_{16}}_{32 \text{ bits}},$$

each of size 32 bits.

- ▶ Padding: Method 2 (to make sure |m| is a multiple of $\ell = 512$)
- ▶ Output: hash of length n = 256 bits



▶ Input: message m of length $0 \le |m| < 2^{64}$ to be divided into t blocks

$$m_1$$
, m_2 , ..., m_t
 512 bits 512 bits 512 bits 512 bits $\leq 2^{64} \approx 16 \times 10^{24} \text{ bits}$

each of length $\ell = 512$ bits.

For each m_i , the $\ell = 512$ bits are further sub-divided into 16 words

$$X_1$$
, X_2 , X_2 , ..., X_{16} , X_{16} ,

each of size 32 bits.

- ▶ Padding: Method 2 (to make sure |m| is a multiple of $\ell = 512$)
- ▶ Output: hash of length n = 256 bits
- Compression function

$$f: \{0,1\}^{\ell+n} \to \{0,1\}^n$$

(64 rounds, each of one step)





For 32-bit inputs u, v, w, x, we define the six functions:

$$f'(u, v, w) = (u \wedge v) \oplus ((\neg u) \wedge w),$$

For 32-bit inputs u, v, w, x, we define the six functions:

$$f'(u, v, w) = (u \wedge v) \oplus ((\neg u) \wedge w),$$

$$g'(u, v, w) = (u \wedge v) \oplus (u \wedge w) \oplus (v \wedge w),$$

For 32-bit inputs u, v, w, x, we define the six functions:

$$f'(u, v, w) = (u \wedge v) \oplus ((\neg u) \wedge w),$$

$$g'(u, v, w) = (u \wedge v) \oplus (u \wedge w) \oplus (v \wedge w),$$

$$\sum_{0}(x) = (x \gg 2) \oplus (x \gg 13) \oplus (x \gg 22),$$

For 32-bit inputs u, v, w, x, we define the six functions:

$$f'(u, v, w) = (u \wedge v) \oplus ((\neg u) \wedge w),$$

$$g'(u, v, w) = (u \wedge v) \oplus (u \wedge w) \oplus (v \wedge w),$$

$$\sum_{0}(x) = (x \gg 2) \oplus (x \gg 13) \oplus (x \gg 22),$$

$$\sum_{1}(x) = (x \gg 6) \oplus (x \gg 11) \oplus (x \gg 25),$$



For 32-bit inputs u, v, w, x, we define the six functions:

$$f'(u, v, w) = (u \land v) \oplus ((\neg u) \land w),$$

$$g'(u, v, w) = (u \land v) \oplus (u \land w) \oplus (v \land w),$$

$$\sum_{0}(x) = (x \gg 2) \oplus (x \gg 13) \oplus (x \gg 22),$$

$$\sum_{1}(x) = (x \gg 6) \oplus (x \gg 11) \oplus (x \gg 25),$$

$$\sigma_{0}(x) = (x \gg 7) \oplus (x \gg 18) \oplus (x \gg 3), \text{ and}$$

For 32-bit inputs u, v, w, x, we define the six functions:

Here,
$$\gg$$
 denotes right rotate and \gg denotes right shift $f'(u, v, w) = (u \land v) \oplus ((\neg u) \land w),$ $g'(u, v, w) = (u \land v) \oplus (u \land w) \oplus (v \land w),$ $\sum_0(x) = (x \gg 2) \oplus (x \gg 13) \oplus (x \gg 22),$ $\sum_1(x) = (x \gg 6) \oplus (x \gg 11) \oplus (x \gg 25),$ $\sigma_0(x) = (x \gg 7) \oplus (x \gg 18) \oplus (x \gg 3),$ and $\sigma_1(x) = (x \gg 17) \oplus (x \gg 19) \oplus (x \gg 10).$



We define 64 constant words: K_0, \ldots, K_{63} .

Each K_i is the first 32 bits of the fractional parts of the cube roots of the first 64 prime numbers.

 K_0 first 32 bits of the fractional part of $\sqrt[3]{2}$,

We define 64 constant words: K_0, \ldots, K_{63} .

Each K_i is the first 32 bits of the fractional parts of the cube roots of the first 64 prime numbers.

 K_0 first 32 bits of the fractional part of $\sqrt[3]{2}$,

 K_1 first 32 bits of the fractional part of $\sqrt[3]{3}$,

We define 64 constant words: K_0, \ldots, K_{63} .

Each K_i is the first 32 bits of the fractional parts of the cube roots of the first 64 prime numbers.

 K_0 first 32 bits of the fractional part of $\sqrt[3]{2}$,

 K_1 first 32 bits of the fractional part of $\sqrt[3]{3}$,

 K_2 first 32 bits of the fractional part of $\sqrt[3]{5}$,



We define 64 constant words: K_0, \ldots, K_{63} .

Each K_i is the first 32 bits of the fractional parts of the cube roots of the first 64 prime numbers.

 K_0 first 32 bits of the fractional part of $\sqrt[3]{2}$,

 K_1 first 32 bits of the fractional part of $\sqrt[3]{3}$,

 K_2 first 32 bits of the fractional part of $\sqrt[3]{5}$,

!



We define 64 constant words: K_0, \dots, K_{63} .

Each K_i is the first 32 bits of the fractional parts of the cube roots of the first 64 prime numbers.

 K_0 first 32 bits of the fractional part of $\sqrt[3]{2}$,

 K_1 first 32 bits of the fractional part of $\sqrt[3]{3}$,

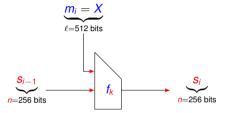
 K_2 first 32 bits of the fractional part of $\sqrt[3]{5}$,

: :

 K_{63} first 32 bits of the fractional part of $\sqrt[3]{p_{63}}$

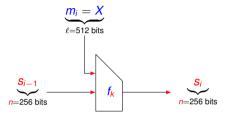


SHA-256: Internal State s_i



The Building Block: A Compression Function $f:\{0,1\}^{\ell+n} \to \{0,1\}^n$

SHA-256: Internal State s_i



The Building Block: A Compression Function $f: \{0,1\}^{\ell+n} \to \{0,1\}^n$

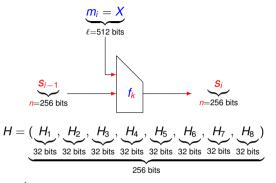
The internal state s_i is a set of 8 values, each of size 32 bits.

$$H = (\underbrace{H_1}_{32 \text{ bits } 32 \text{ bits } 32$$

This also corresponds to the 256 bits of input key s_0 / output s_t



SHA-256: Initial State s₀



For the initial state s_0 , we assign:

$$\begin{array}{lll} H_1 \leftarrow 0x6A09E667, & H_2 \leftarrow 0xBB67AE85, \\ H_3 \leftarrow 0x3C6EF372, & H_4 \leftarrow 0xA54FF53A, \\ H_5 \leftarrow 0x510E527F, & H_6 \leftarrow 0x9B05688C, \\ H_7 \leftarrow 0x1F83D9AB, \text{ and} & H_8 \leftarrow 0x5BE0CD19. \end{array}$$





SHA-256: The Compression Function

SHA-256 compression function $f(X||s_{i-1})$

Initialisation:
$$(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8) \leftarrow (H_1, H_2, H_3, H_4, H_5, H_6, H_7, H_8)$$

Expansion: for
$$j = 17$$
 to 64 do

$$X_i \leftarrow (\sigma_1(X_{i-2}) + X_{i-7} + \sigma_0(X_{i-15}) + X_{i-16})$$

Rounds: for
$$j = 1$$
 to 64 do

$$t_{1} \leftarrow (Y_{8} + \sum_{1}(Y_{5}) + f'(Y_{5}, Y_{6}, Y_{7}) + K_{i} + X_{i})$$

$$t_{2} \leftarrow (\sum_{0}(Y_{1}) + g'(Y_{1}, Y_{2}, Y_{3}))$$

$$Y_{1} \leftarrow (t_{1} + t_{2})$$

$$Y_{2}, Y_{3}, Y_{4} \leftarrow Y_{1}, Y_{2}, Y_{3}$$

$$Y_{5} \leftarrow Y_{4} + t_{1}$$

$$Y_{6}, Y_{7}, Y_{8} \leftarrow Y_{5}, Y_{6}, Y_{7}$$
for $k = 1$ to 8
$$H_{k} \leftarrow Y_{k} + H_{k}$$

$$S_{i} \leftarrow (H_{1}, H_{2}, H_{3}, H_{4}, H_{5}, H_{6}, H_{7}, H_{8})$$

Popular Hash Functions H(n)

Check this link: https://www.browserling.com/tools/all-hashes

Check the following SHA-256 hashes:

SHA256("WISDOM IS NOWHERE") =

"7db96cca6d0160496161297d522425ffab681f25c6bd57c85bf976a8b7d7372d"

SHA256("WISDOM IS NOW HERE") =

"eae 0 bc 219689024 e 359 b 23 a 1 a 3 e 81 be 946423 be 9a7f 2332 c 430 f 7a 17c 1a 315 e 7".

Note: Each character in a hash is the hexadecimal encoding of 4 binary digits.

Hex	Binary	Hex	Binary	Hex	Binary	Hex	Binary
0	0000	4	0100	8	1000	С	1100
1	0001	5	0101	9	1001	d	1101
2	0010	6	0110	a	1010	е	1110
3	0011	7	0111	b	1011	f	1111

Note: A "slight" change in the input, completely changes the hash value!

Very useful for implementing integrity checks!







Thank you for your kind attention!



Easy to search by name (sorted, hence binary search); Hard to search by number (unsorted, hence exhaustive search)

Budi Arief	07932731720
Adam Back	07896572341
David Barnes	07723812346
Mark Batty	07983184516
Laura Bocchi	07988762343
Howard Bowman	07981722451
Janet Carter	07986662211
David Chadwick	07766923565
Jacqueline Chetty	07898234553
Olaf Chitil	07799012748
Theodosios Dimitrakos	07978343434
Sally Fincher	07898347543
Virginia Franqueira	07711524511
Radu Grigore	07776351939
Marek Grzes	07928672498
Julio Hernandez-Castro	07717210073
Tim Hopkins	07792378464
Ozgur Kafali	07989634247
Stefan Kahrs	07798127475
Stephen Kell	07987234122
Andy King	07898374335
Julien Lange	07912685674
Rogerio de Lemos	07973542352
Shujun Li	07713467929
Stefan Marr	07987641123
Jason Nurse	07777712981
Dominic Orchard	07981279642
Carlos Perez Delgado	07772129875
Simon Thompson	07981237412
Charles Xavier	07927862449





Easy to search by name (sorted, hence binary search); Hard to search by number (unsorted, hence exhaustive search)

Budi Arief	07932731720
Adam Back	07896572341
David Barnes	07723812346
Mark Batty	07983184516
Laura Bocchi	07988762343
Howard Bowman	07981722451
Janet Carter	07986662211
David Chadwick	07766923565
Jacqueline Chetty	07898234553
Olaf Chitil	07799012748
Theodosios Dimitrakos	07978343434
Sally Fincher	07898347543
Virginia Franqueira	07711524511
Radu Grigore	07776351939
Marek Grzes	07928672498
Julio Hernandez-Castro	07717210073
Tim Hopkins	07792378464
Ozgur Kafali	07989634247
Stefan Kahrs	07798127475
Stephen Kell	07987234122
Andy King	07898374335
Julien Lange	07912685674
Rogerio de Lemos	07973542352
Shujun Li	07713467929
Stefan Marr	07987641123
Jason Nurse	07777712981
Dominic Orchard	07981279642
Carlos Perez Delgado	07772129875
Simon Thompson	07981237412
Charles Xavier	07927862449

Phone number of Jason Nurse?





Easy to search by name (sorted, hence binary search); Hard to search by number (unsorted, hence exhaustive search)

Budi Arief	07932731720
Adam Back	07896572341
David Barnes	07723812346
Mark Batty	07983184516
Laura Bocchi	07988762343
Howard Bowman	07981722451
Janet Carter	07986662211
David Chadwick	07766923565
Jacqueline Chetty	07898234553
Olaf Chitil	07799012748
Theodosios Dimitrakos	07978343434
Sally Fincher	07898347543
Virginia Franqueira	07711524511
Radu Grigore	07776351939
Marek Grzes	07928672498
Julio Hernandez-Castro	07717210073
Tim Hopkins	07792378464
Ozgur Kafali	07989634247
Stefan Kahrs	07798127475
Stephen Kell	07987234122
Andy King	07898374335
Julien Lange	07912685674
Rogerio de Lemos	07973542352
Shujun Li	07713467929
Stefan Marr	07987641123
Jason Nurse	07777712981
Dominic Orchard	07981279642
Carlos Perez Delgado	07772129875
Simon Thompson	07981237412
Charles Xavier	07927862449

Phone number of Jason Nurse?





Easy to search by name (sorted, hence binary search); Hard to search by number (unsorted, hence exhaustive search)

ortou, nonce contactive	004.0.1)
Budi Arief	07932731720
Adam Back	07896572341
David Barnes	07723812346
Mark Batty	07983184516
Laura Bocchi	07988762343
Howard Bowman	07981722451
Janet Carter	07986662211
David Chadwick	07766923565
Jacqueline Chetty	07898234553
Olaf Chitil	07799012748
Theodosios Dimitrakos	07978343434
Sally Fincher	07898347543
Virginia Franqueira	07711524511
Radu Grigore	07776351939
Marek Grzes	07928672498
Julio Hernandez-Castro	07717210073
Tim Hopkins	07792378464
Ozgur Kafali	07989634247
Stefan Kahrs	07798127475
Stephen Kell	07987234122
Andy King	07898374335
Julien Lange	07912685674
Rogerio de Lemos	07973542352
Shujun Li	07713467929
Stefan Marr	07987641123
Jason Nurse	07777712981
Dominic Orchard	07981279642
Carlos Perez Delgado	07772129875
Simon Thompson	07981237412
Charles Xavier	07927862449

Phone number of Jason Nurse?

Easy!





Easy to search by name (sorted, hence binary search);

Hard to search by number (unsorted, hence exhaustive search)

Budi Arief	07932731720
Adam Back	07896572341
David Barnes	07723812346
Mark Batty	07983184516
Laura Bocchi	07988762343
Howard Bowman	07981722451
Janet Carter	07986662211
David Chadwick	07766923565
Jacqueline Chetty	07898234553
Olaf Chitil	07799012748
Theodosios Dimitrakos	07978343434
Sally Fincher	07898347543
Virginia Franqueira	07711524511
Radu Grigore	07776351939
Marek Grzes	07928672498
Julio Hernandez-Castro	07717210073
Tim Hopkins	07792378464
Ozgur Kafali	07989634247
Stefan Kahrs	07798127475
Stephen Kell	07987234122
Andy King	07898374335
Julien Lange	07912685674
Rogerio de Lemos	07973542352
Shujun Li	07713467929
Stefan Marr	07987641123
Jason Nurse	07777712981
Dominic Orchard	07981279642
Carlos Perez Delgado	07772129875
Simon Thompson	07981237412
Charles Xavier	07927862449

Phone number of Jason Nurse?

Easy!

Where is 07973542352?





Easy to search by name (sorted, hence binary search);

Hard to search by number (unsorted, hence exhaustive search)

Budi Arief	07932731720
Adam Back	07896572341
David Barnes	07723812346
Mark Batty	07983184516
Laura Bocchi	07988762343
Howard Bowman	07981722451
Janet Carter	07986662211
David Chadwick	07766923565
Jacqueline Chetty	07898234553
Olaf Chitil	07799012748
Theodosios Dimitrakos	07978343434
Sally Fincher	07898347543
Virginia Franqueira	07711524511
Radu Grigore	07776351939
Marek Grzes	07928672498
Julio Hernandez-Castro	07717210073
Tim Hopkins	07792378464
Ozgur Kafali	07989634247
Stefan Kahrs	07798127475
Stephen Kell	07987234122
Andy King	07898374335
Julien Lange	07912685674
Rogerio de Lemos	07973542352
Shujun Li	07713467929
Stefan Marr	07987641123
Jason Nurse	07777712981
Dominic Orchard	07981279642
Carlos Perez Delgado	07772129875
Simon Thompson	07981237412
Charles Xavier	07927862449

Phone number of Jason Nurse?

Easy!

Where is 07973542352?





Easy to search by name (sorted, hence binary search);

Hard to search by number (unsorted, hence exhaustive search)

Budi Arief	07932731720
Adam Back	07896572341
David Barnes	07723812346
Mark Batty	07983184516
Laura Bocchi	07988762343
Howard Bowman	07981722451
Janet Carter	07986662211
David Chadwick	07766923565
Jacqueline Chetty	07898234553
Olaf Chitil	07799012748
Theodosios Dimitrakos	07978343434
Sally Fincher	07898347543
Virginia Franqueira	07711524511
Radu Grigore	07776351939
Marek Grzes	07928672498
Julio Hernandez-Castro	07717210073
Tim Hopkins	07792378464
Ozgur Kafali	07989634247
Stefan Kahrs	07798127475
Stephen Kell	07987234122
Andy King	07898374335
Julien Lange	07912685674
Rogerio de Lemos	07973542352
Shujun Li	07713467929
Stefan Marr	07987641123
Jason Nurse	07777712981
Dominic Orchard	07981279642
Carlos Perez Delgado	07772129875
Simon Thompson	07981237412
Charles Xavier	07927862449

Phone number of Jason Nurse?

Easy!

Where is 07973542352?

Hard!



