Hashes

Lesson Introduction

- The birthday paradox and length of hash
- Secure hash function
- HMAC

Hash Functions

- Compute message digest of data of any size
- Fixed length output: 128-512 bits
- Easy to compute $H(m)$
- Given $H(m)$, no easy way to find m
  - *One-way function*
- Given m_1, it is computationally infeasible to find $m_2 \neq m_1$ s.t. $H(m_2) = H(m_1)$
  - Weak collision resistant
- Computationally infeasible to find $m_1 \neq m_2$ s.t. $H(m_1) = H(m_2)$
  - Strong collision resistant

Hash Functions

- Compute message digest of data of any size
- Fixed length output: 128-512 bits
- Easy to compute $H(m)$
- Given $H(m)$, no easy way to find m
  - *One-way function*
- Given m_1, it is computationally infeasible to find $m_2 \neq m_1$ s.t. $H(m_2) = H(m_1)$
  - Weak collision resistant
- Computationally infeasible to find $m_1 \neq m_2$ s.t. $H(m_1) = H(m_2)$
  - Strong collision resistant

Requirements for a practical application of a hash function

Hash Functions

- Compute message digest of data of any size
- Fixed length output: 128-512 bits
- Easy to compute $H(m)$
- Given $H(m)$, no easy way to find m
  - *One-way function*

The one way property

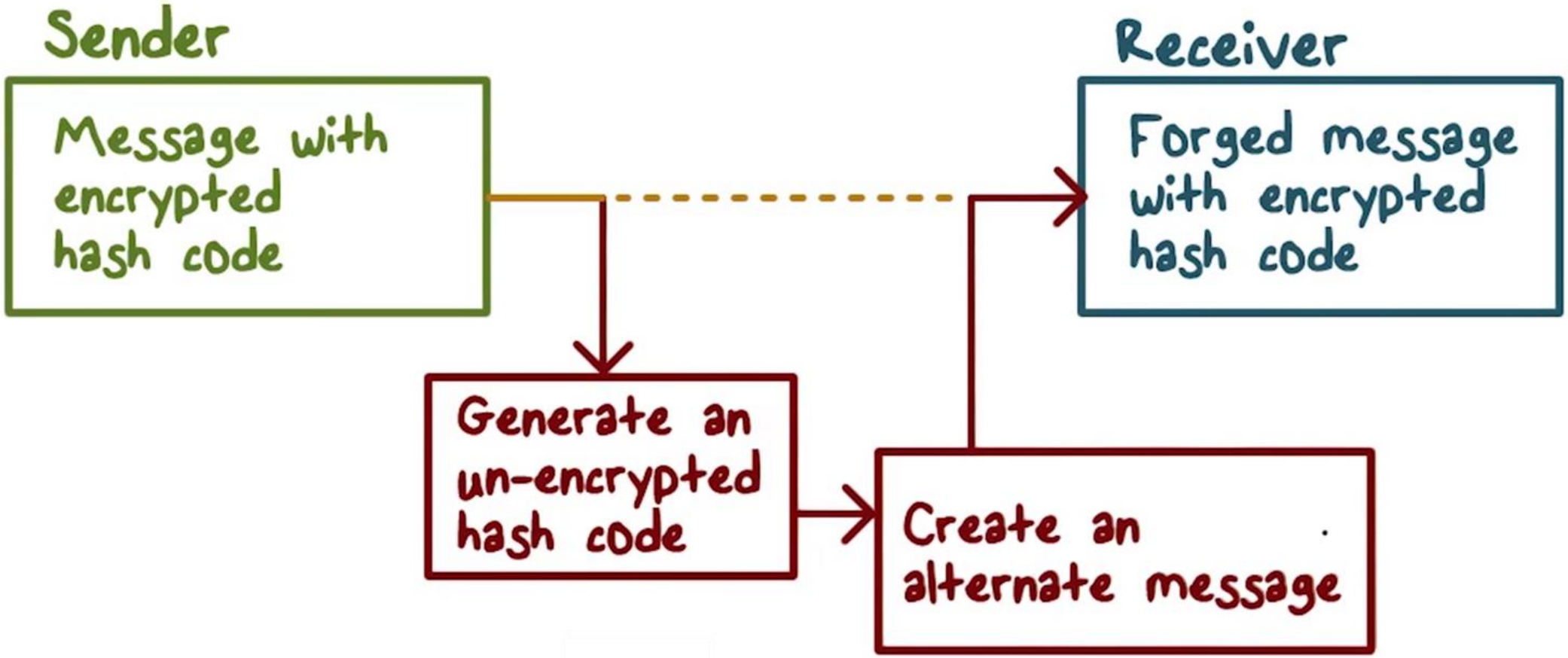
- Given m_1, it is computationally infeasible to find $m_2 \neq m_1$ s.t. $H(m_2) = H(m_1)$
  - Weak collision resistant
- Computationally infeasible to find $m_1 \neq m_2$ s.t. $H(m_1) = H(m_2)$
  - Strong collision resistant

Hash Functions

- Compute message digest of data of any size
- Fixed length output: 128-512 bits
- Easy to compute $H(m)$
- Given $H(m)$, no easy way to find m
  - *One-way function*
- Given m_1, it is computationally infeasible to find $m_2 \neq m_1$ s.t. $H(m_2) = H(m_1)$
  - Weak collision resistant
- Computationally infeasible to find $m_1 \neq m_2$ s.t. $H(m_1) = H(m_2)$
  - Strong collision resistant

Hash functions are unique to each message

Hash Functions

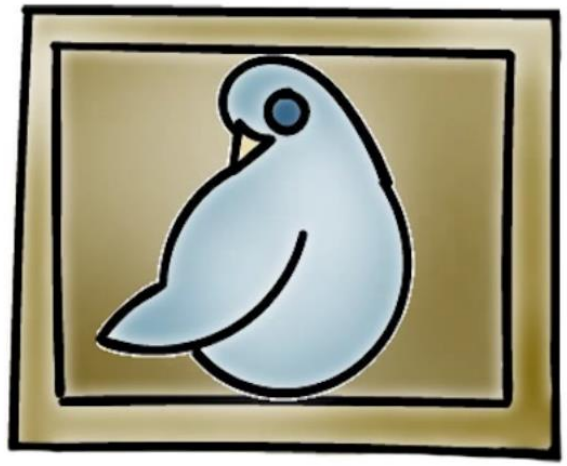


Hash Functions

- Compute message digest of data of any size
- Fixed length output: 128-512 bits
- Easy to compute $H(m)$
- Given $H(m)$, no easy way to find m
  - *One-way function*
- Given m_1, it is computationally infeasible to find $m_2 \neq m_1$ s.t. $H(m_2) = H(m_1)$
  - Weak collision resistant
- Computationally infeasible to find $m_1 \neq m_2$ s.t. $H(m_1) = H(m_2)$
  - Strong collision resistant

Hash Functions

Hash Function Weaknesses

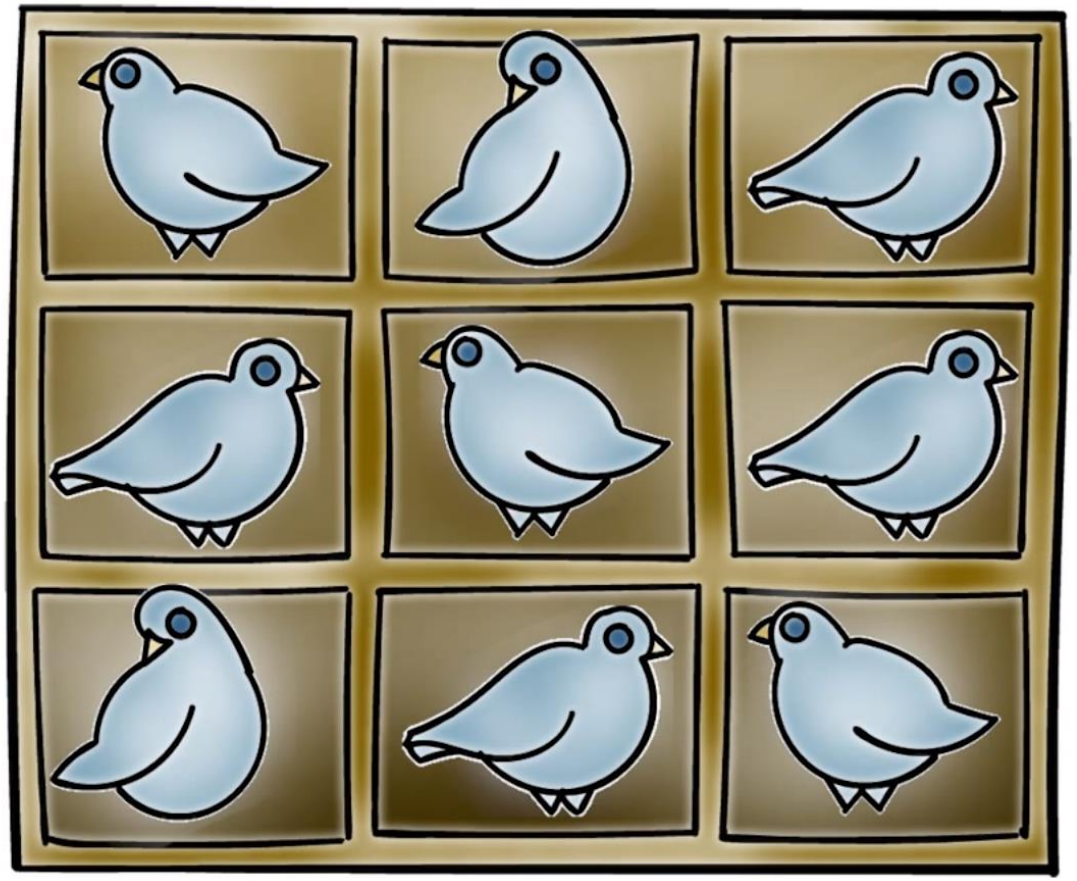


Pigeonhole Principle

The Birthday Paradox

Hash Function Weaknesses

Pigeonhole Principle



n = number of pigeons
m = number of holes

n = m There is one pigeon per hole

n > m Then at least one hole must have more than one pigeon

Hash Function Weaknesses

The Birthday Paradox

How many people do you need in a room before you have a **greater than 50% chance** that two of them will have the same birthday?

Assume 365 birthdays (our containers)

% chance that two people in the room have the same birthday:

100% requires 366 people (the pigeonhole principle)

Hash Function Weaknesses

The Birthday Paradox

- **Compute probability of different birthdays**
- Random sample of n people (birthdays) taken from *k* (365) days
- k^n samples with replacement
- $(k)_n = k(k-1)\ldots(k-n+1)$ sample without replacement
- **Probability of repetition**:
  - $p = 1 - (k)_n/k^n \approx n(n-1)/2k = 0.5$ if $n=\sqrt{k}$

Hash Function Weaknesses

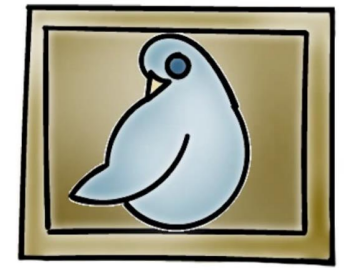
The Birthday Paradox

$1-(k)_n/k^n$= **the probability that a pair share the same birthday**

If k = 365, n = 19

If there are **19 people in a room**, there is a good chance that **two of them** share the same birthday!

Hash Function Weaknesses



Hash Functions:

- There are many more ‘pigeons’ than ‘pigeonholes’

- Many inputs will be mapped to the same output. That is, ***many input messages will have the same hash.***

Conclusion: The longer the length of the hash, the fewer collisions.

Determining Hash Length

Hash Length	Possible # of hash values
1	2
64	2^{32}

Hash Size Quiz

Choose the correct answer:

If the length of hash is 128 bits, then how many messages does an attack need to search in order to find two that share the same hash?

- ☐ 128
- ☐ 2^{127}
- ☐ 2^{128}
- ☐ 2^{64}

Secure Hash Algorithm

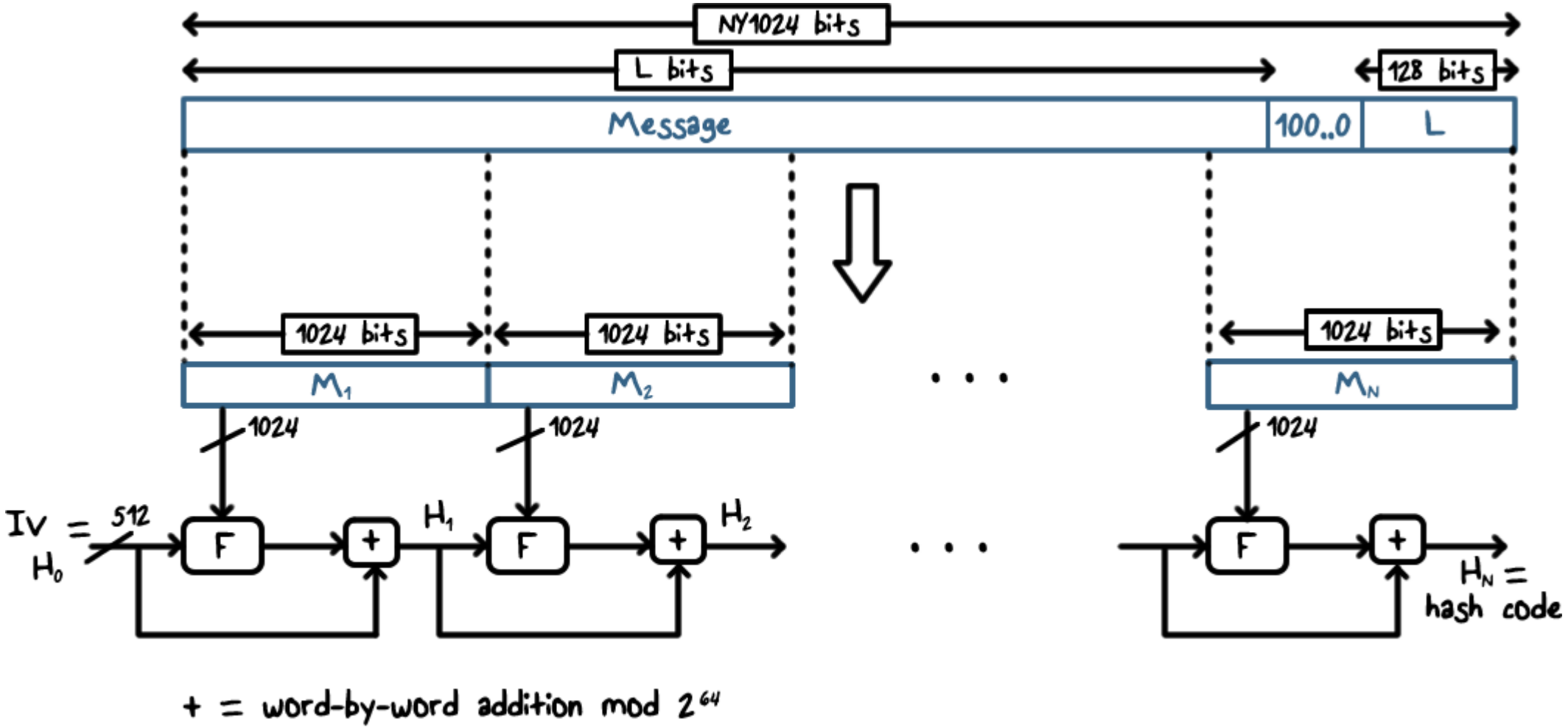
- **Developed by NIST**, specified in the Secure Hash Standard, originally 1993
- Revised as SHA-1 in 1995
  - 160 bit hash
- **NIST specified SHA2 algorithms in 2002**
  - Hash value lengths of 256, 384, and 512
  - Similar to SHA-1

Comparison of SHA Parameters

	SHA-1	SHA-256	SHA-384	SHA-512
Message digest size	160	256	384	512
Message size	$<2^{64}$	$<2^{64}$	$<2^{128}$	$<2^{128}$
Block size	512	512	1024	1024
Word size	32	32	64	64
Number of steps	80	80	80	80
Security	80	128	192	256

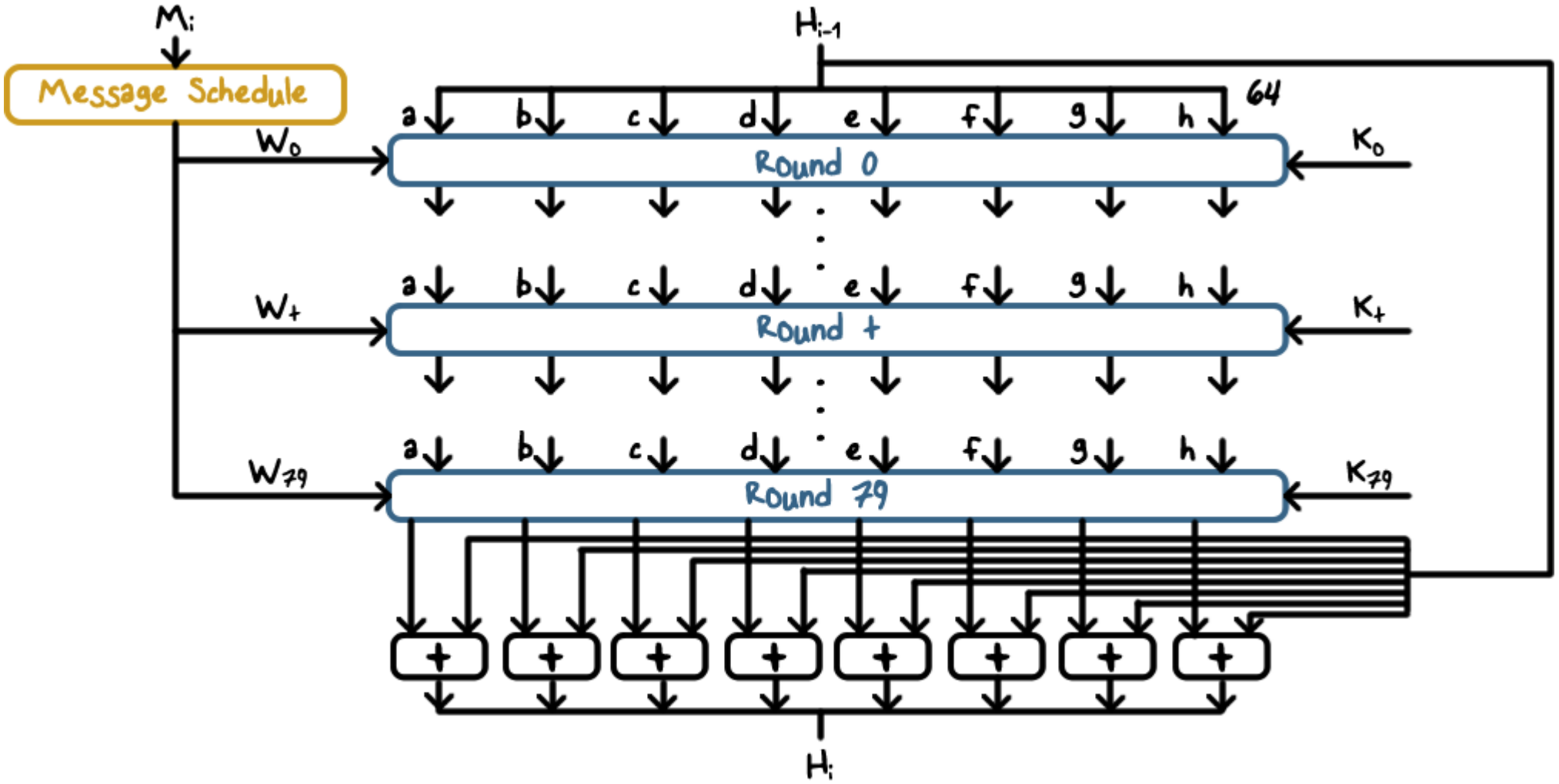
Notes:
1. All sizes are measured in bits.
2. Security refers to the fact that a birthday attack on a message digest of size n produces a collision with a work factor of approximately $2^{n/2}$.

Message Processing



Message Digest Generation Using SHA-512

Message Processing

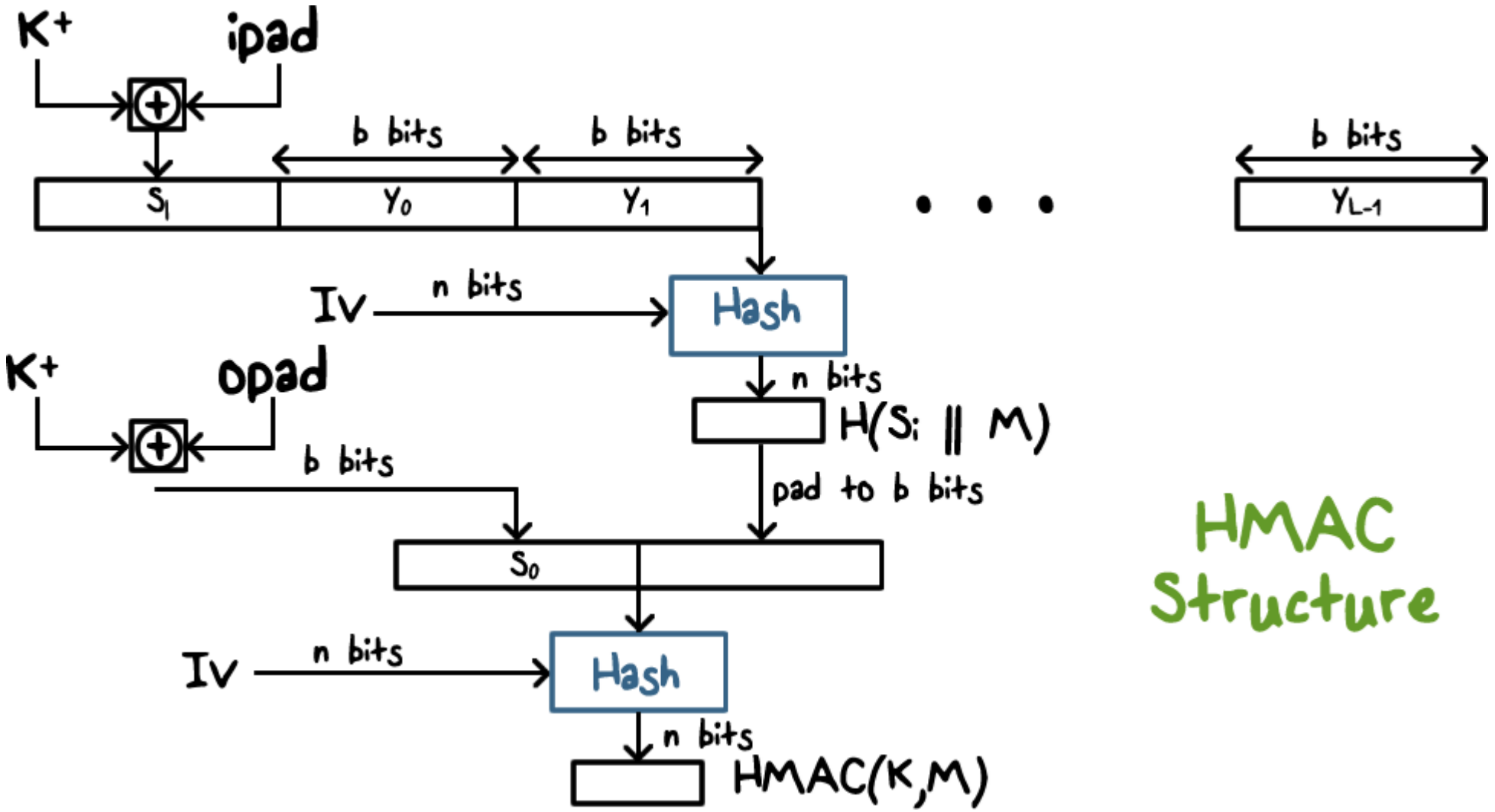


SHA-512 Processing of a Single 1024-Bit Block

Hash Based Message Authentication

- **Cryptographic hash functions generally execute faster**
- Library code is widely available
- SHA-1 was not designed for use as a MAC because it does not rely on a secret key
- Issued as RFC2014
- **Has been chosen as the mandatory-to-implement MAC for IP security**
- Used in other Internet protocols such as Transport Layer Security (TLS)

Hash Based Message Authentication



HMAC Structure

HMAC Security

- Security **depends on the cryptographic strength** of the underlying hash function

- It's much **harder to launch successful collision attacks on HMAC** because of secret key

Hash Function Quiz

Check the statements that are True:

- ☐ The one-way hash function is important not only in message authentication but also in digital signatures
- ☐ SHA processes the input one block at a time but each block goes through the same processing
- ☐ HMAC is secure provided that the embedded hash function has good cryptographic strengths such as one-way and collision resistant

Hashes

Lesson Summary

- Hash length should be at least 128
  - 2^(64) message to find collision
- SHA1: 160-bit hash; SHA2: 256/384/512-bit hash
  - Message processed/hashed block by block, result to next
- HMAC: hash the message with a secret key
  - Message authentication