# Funzioni di Hash

**Prof. Francesco Bergadano**

**Dipartimento di Informatica**
**Università degli Studi di Torino**

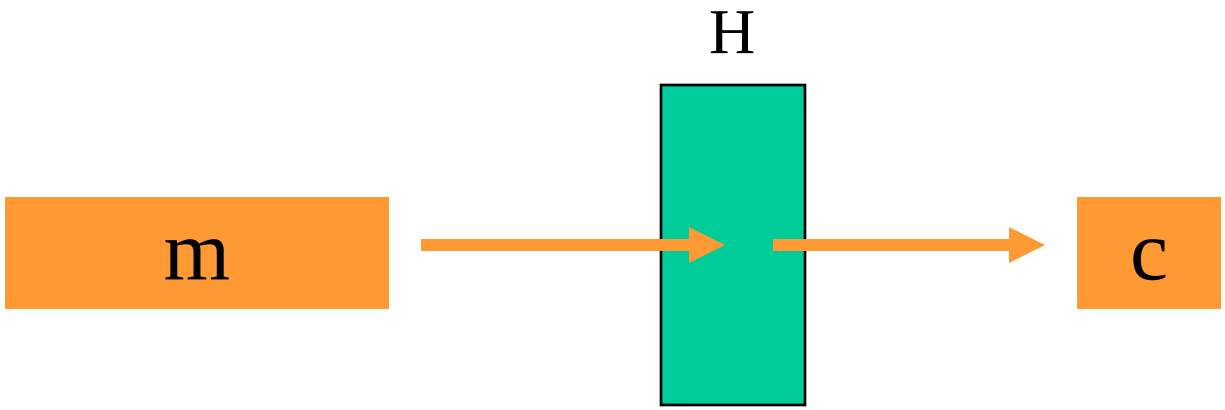
# Funzione di hash H

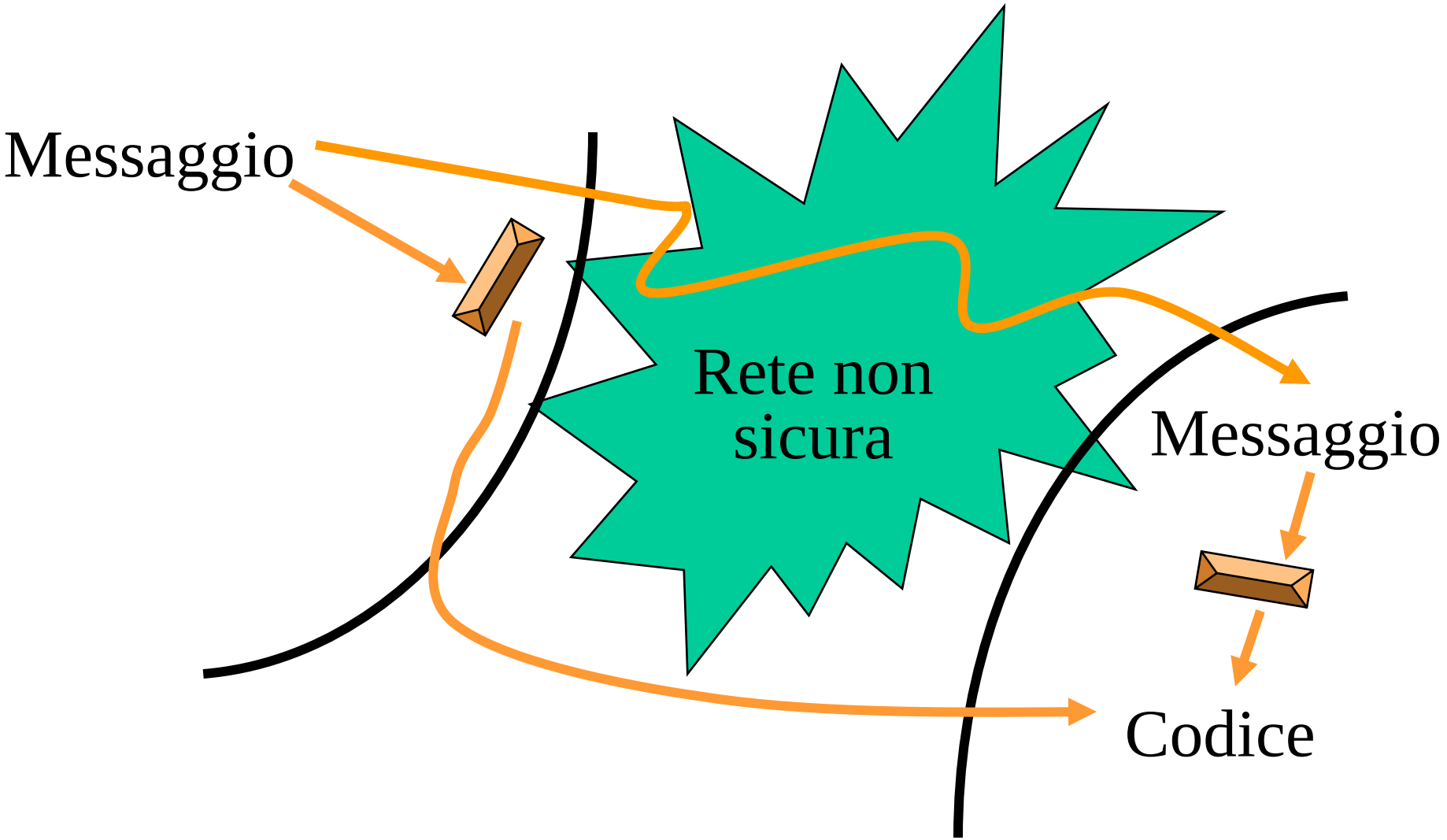
**Traforma un messaggio m di lunghezza variabile in un codice c di lunghezza fissa**

**H(m) = c**

***Essenziali in molte applicazioni, e utilizzate per generare firme elettroniche.***

# Funzione di hash H






Può essere usata per semplici forme di autenticazione

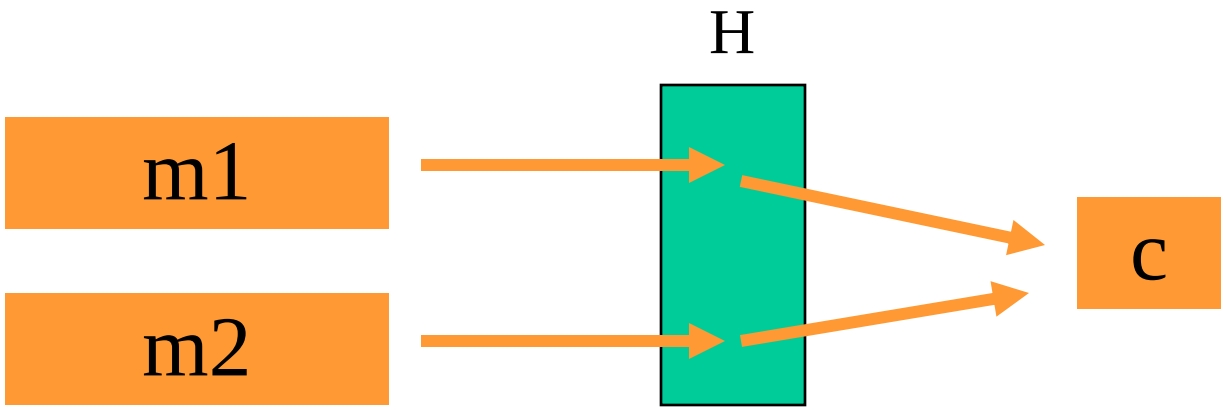
# Funzione di hash H

**Siccome, in media |m| > |c|,**
**può succedere che**

**H(m1) = c**
**H(m2) = c**

**<m1,m2> è una collisione per H**

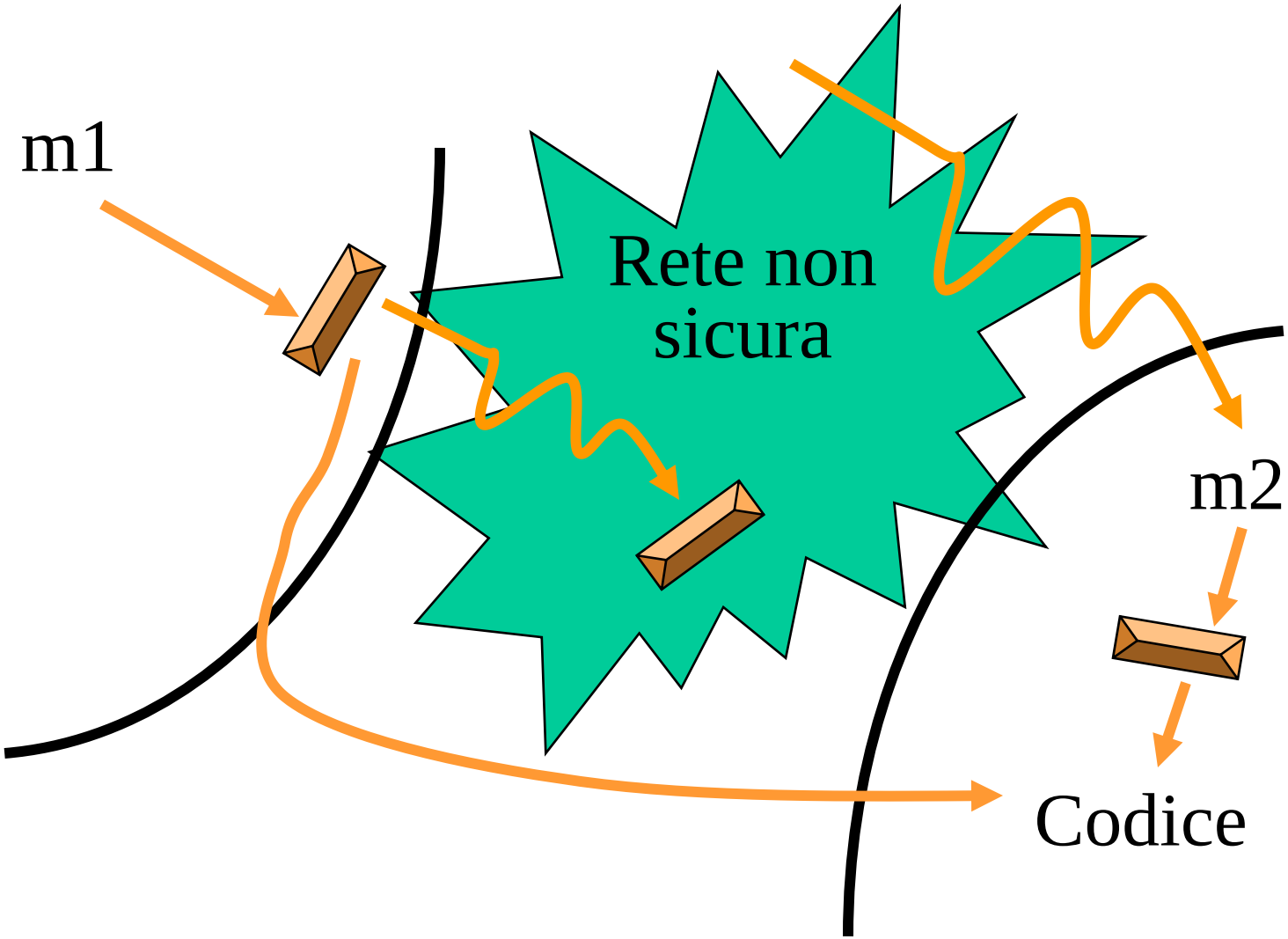
# Collisioni di una funzione di hash



# Collisioni = Problemi

**H(m1) = c**

**H(m2) = c**




Le collisioni rendono possibile la falsificazione di messaggi

# Se è facile calcolare $H^{-1}$ sarà facile ottenere collisioni

**$H^{-1}(c) = \{m1,m2\}$**

# Desiderata per una funzione di hash

- **Non invertibile** o a una via (one-way): dato c, è difficile trovare m tale che $H^{-1}(c)$ = insieme dei messaggi m t.c. H(m)=c
- **Fortemente non invertibile**: dato m1, è difficile trovare m2 tale che H(m1) = H(m2)
- **Resistente alle collisioni** (collision resistant): è difficile trovare m1 e m2 tali che H(m1) = H(m2)

# Desiderata per una funzione di hash

**Non invertibile**

**Fortemente non invertibile**

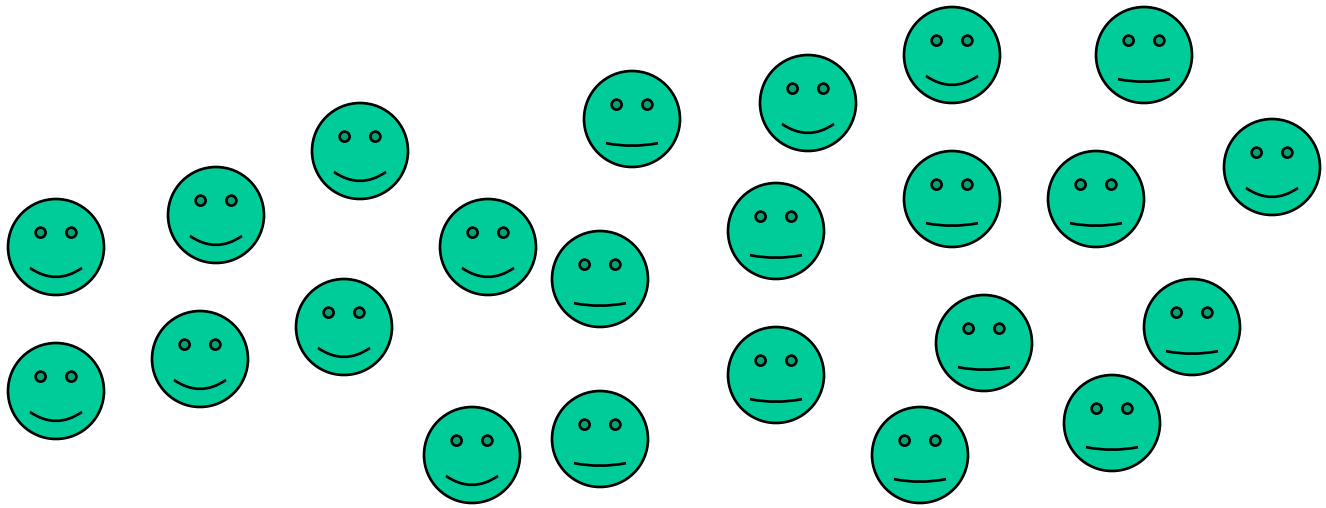
**Resistente alle collisioni**

# Attacchi ‘del compleanno’ ad una funzione di hash H (birthday attacks)

- **Sono metodi per generare collisioni <m1,m2> di H**
- **Permettono di scegliere m1 ed m2 in modo che m1∈M1 e m2∈M2, dove M1 e M2 sono insiemi di messaggi predeterminati**

# Il paradosso del compleanno

**La probabilità che in un gruppo di 23 persone due abbiano lo stesso compleanno è maggiore di 1/2.**



# Il paradosso del compleanno

**La probabilità che in un gruppo di 23 persone due abbiano lo stesso compleanno è maggiore di 1/2.**

P(nessun compleanno in comune) =

$= (365*364*\ldots*(365-22))/365^{23} = 0.4927$

# In generale:

P(almeno una ripetizione in un insieme di k elementi scelti tra n) = P(n,k) =

$= 1 - (n*(n-1)*\ldots*(n-k+1))/n^k >$

$> 1 - e^{-k*(k-1)/2n} \approx 1 - e^{-k^2/2n}$

$P(n,k) > 0.5$ per $k > 1.18*\sqrt{n}$

*Nel caso del compleanno $22.54 = 1.18*\sqrt{365}$.*

$1-(n*(n-1)*\ldots*(n-k+1))/ n^k =$
$1-[1*(n-1)/n * \ldots * (n-k+1)/n] =$
$1-[(1-1/n)*\ldots*(1-(k-1)/n)] >$
$1-[(e^{-1/n})*\ldots*(e^{-(k-1)/n})] > 1-e^{-k(k-1)/2n}$

P(almeno una ri
di k elementi scelti tra n) = P(n
$= 1- (n*(n-1)*\ldots*(n-k+1))/n^k >$
$> 1 - e^{-k*(k-1)/2n} \approx 1- e^{-k^2/2n}$

$P(n,k) > 0.5$ per $k > 1.18*\sqrt{n}$

*Nel caso del compleanno $22.54 = 1.18*\sqrt{365}$.*

$1-(n*(n-1)*\ldots*(n-k+1))/ n^k$

$1-[(n-1)/n * \ldots * (n-k+1)/n]$

$1-[(1-1/n)*\ldots*(1-(k-1)/n)] >$

$1-[(e^{-1/n})*\ldots*(e^{-(k-1)/n})] > 1-e^{-k(k-1)/2n}$

$e^{-x} > 1-x$

P(almeno una rip

di k elementi scelti tra n) = P(n

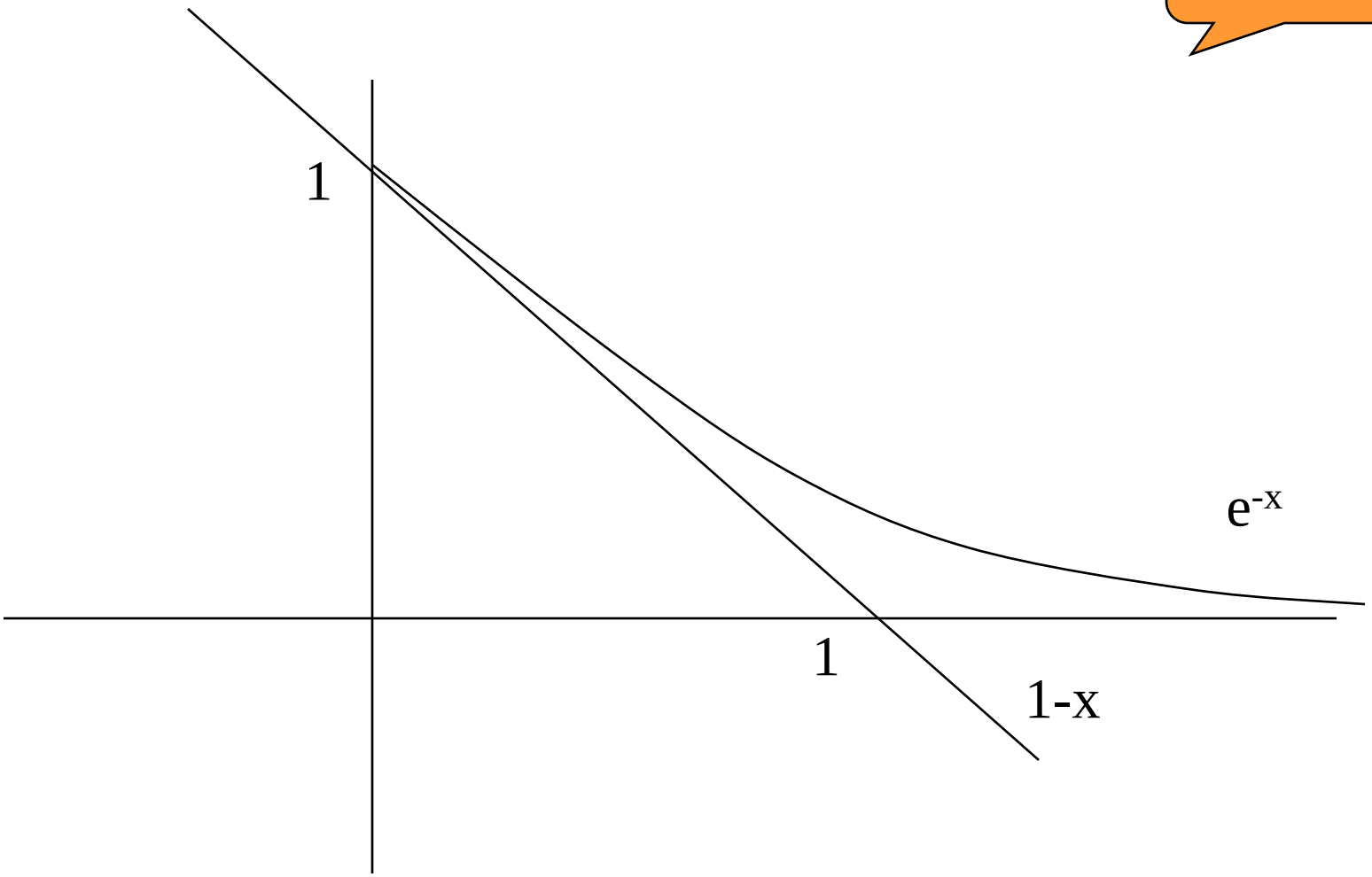
$= 1- (n*(n-1)*\ldots*(n-k+1))/n^k >$

$> 1 - e^{-k*(k-1)/2n} \approx 1- e^{-k^2/2n}$

$P(n,k) > 0.5$ per $k > 1.18*\sqrt{n}$

*Nel caso del compleanno $22.54 = 1.18*\sqrt{365}$.*

$e^{-x} > 1-x$



$1-(n*(n-1)*\ldots*(n-k+1))/\ n^k =$
$1-[(n-1)/n\ *\ \ldots\ *\ (n-k+1)/n] =$
$1-[(1-1/n)*\ldots*(1-(k-1)/n)] >$
$1-[(e^{-1/n})*\ldots*(e^{-(k-1)/n})] > 1-e^{-k(k-1)/2n}$

$(e^{-1/n})*\ldots*(e^{-(k-1)/n}) =$
$e^{-[1\ +\ \ldots\ +\ (k-1)]/n}$

$1+2+\ldots+k-1 = k(k-1)/2$

]

elti tra n) = P(n

..*(n-k+1))/$n^k$ >

$> 1 - e^{-k*(k-1)/2n} \approx 1- e^{-k^2/2n}$

$P(n,k) > 0.5$ per $k > 1.18*\sqrt{n}$

*Nel caso del compleanno $22.54 = 1.18*\sqrt{365}$.*

# In generale:

P(almeno una ripetizione in un insieme di k elementi scelti tra n) = P(n,k) =

$= 1 - (n*(n-1)*\ldots*(n-k+1))/n^k >$

$> 1 - e^{-k*(k-1)/2n} \approx 1 - e^{-k^2/2n}$

$P(n,k) > 0.5$ per $k > 1.18*\sqrt{n}$

*Nel caso del compleanno $22.54 = 1.18*\sqrt{365}$.*

$P(n,k) > 0.5$

$1 - e^{-k^2/2n} > 0.5$

$- e^{-k^2/2n} > - 0.5$

$e^{-k^2/2n} < 1/2$

$e^{k^2/2n} > 2$

$\ln(2) < k^2/2n$

$k > \sqrt{(2\ln(2)n)}$

$k > 1.18*\sqrt{n}$

# generale:

tizione in un insieme

i tra n) = P(n,k) =

(n-k+1))/$n^k$ >

> 1 - $e^{-k*(k-1)}$ ≥ 1- $e^{-k^2/2n}$

$P(n,k) > 0.5$ per $k > 1.18*\sqrt{n}$

*Nel caso del compleanno $22.54 = 1.18*\sqrt{365}$.*

# Conseguenze per funzioni di hash

$P(n,k) > 0.5$ per $k > 1.18*\sqrt{n} \approx \sqrt{n}$

*k = numero totale di messaggi =* $2^m$

*n = numero totale di codici di c bit =* $2^c$

P(n,k) = P(collisione) > 0.5 per

$k > \sqrt{n}$, ovvero per $2^m > \sqrt{2^c}$, ovvero

$2^m > 2^{c/2}$ quindi per $m > c/2$.

# Conseguenze per funzioni di hash

*$2^m$ messaggi, codici (valori di h) di c bit*

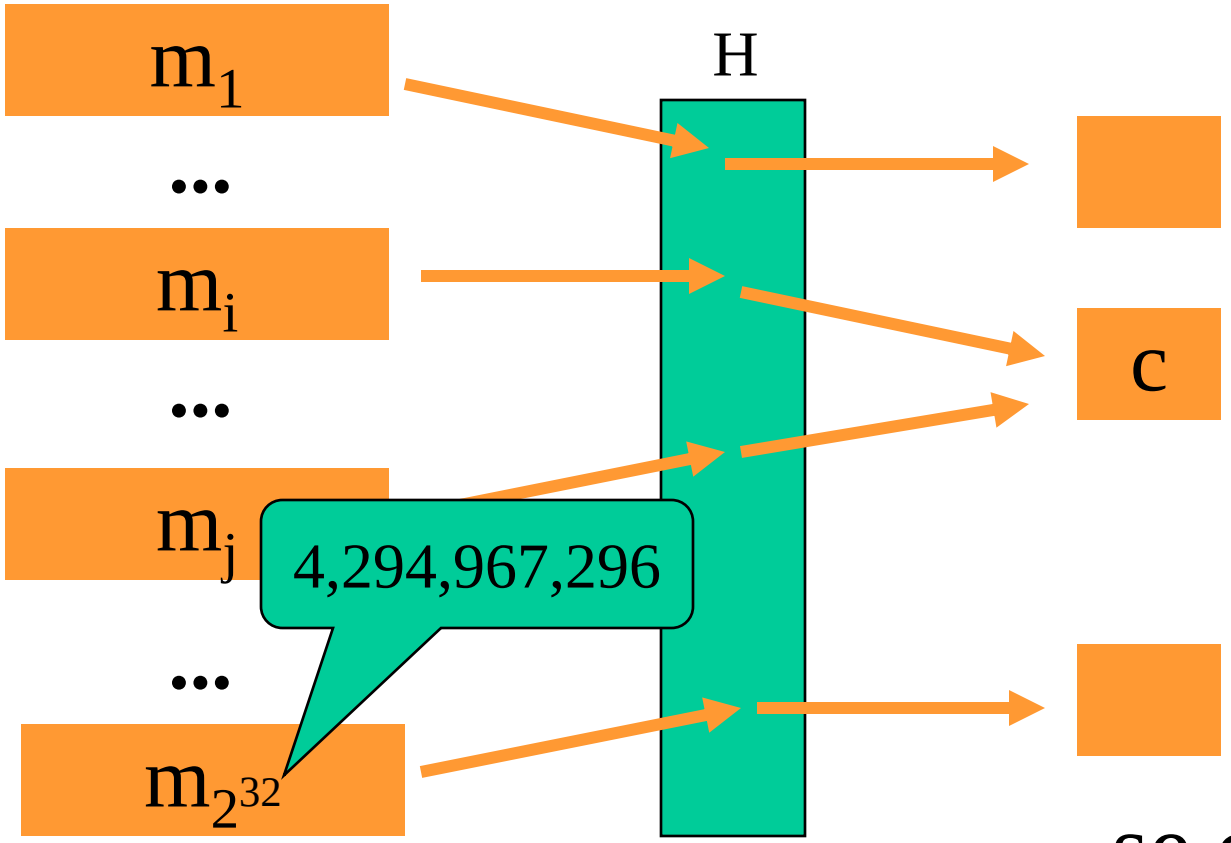
P(collisione) > 0.5 per m > c/2.

Per generare una collisione,

per codici di 64 bit,

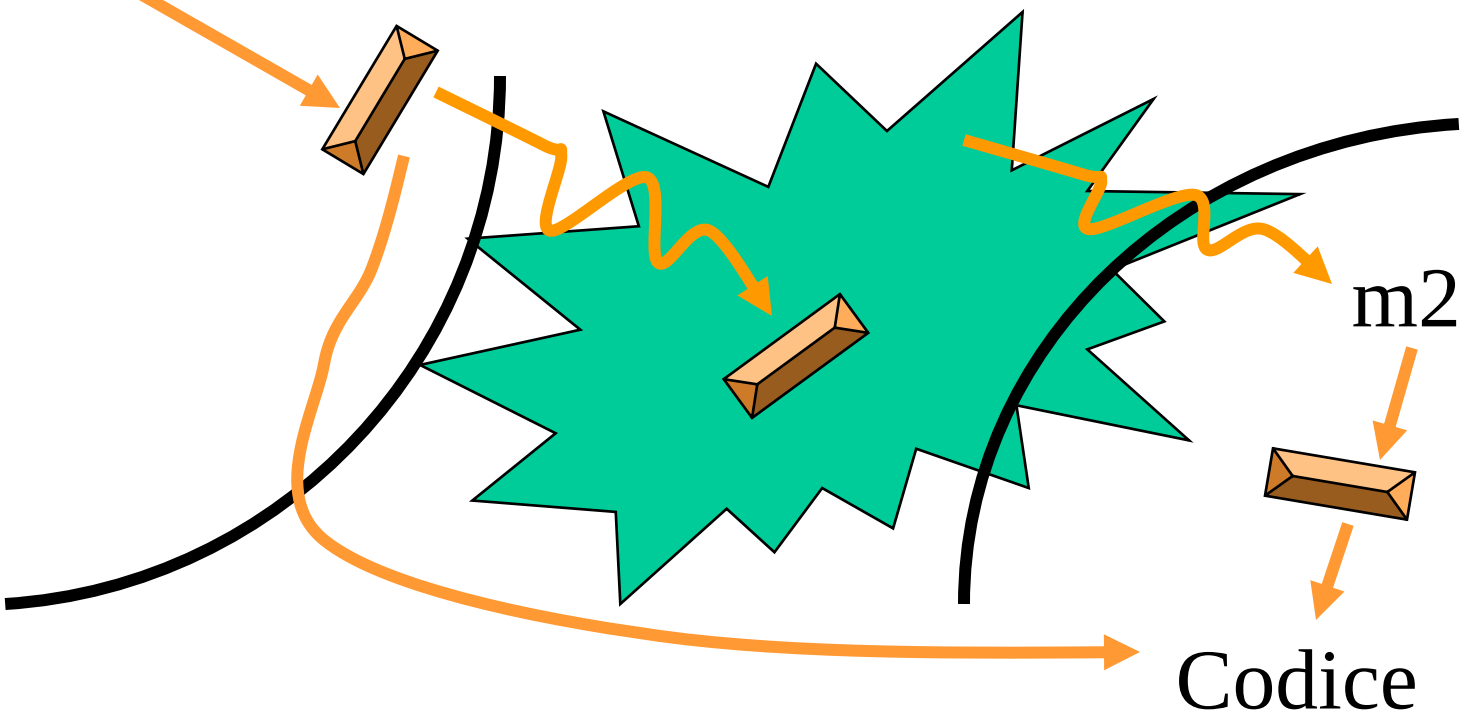
basta provare $2^{32}$ messaggi.

# Con probabilità > 0.5, se c di 64 bit




se c di 64 bit

Ma in questo caso serve
una collisione specifica!




m1 accettabile per mittente
m2 falsificazione desiderata

# Allora il problema è leggermente diverso:

P(almeno un elemento comune in due insiemi di k elementi scelti tra n) = P'(n,k) =

$= 1- ((1- 1/n)^k)^k > 1- ((e^{-1/n})^k)^k = 1-e^{-k^2/n}$.

$P'(n,k) > 0.5$ per $k > 0.83*\sqrt{n}$

X = {x1, x2, …, xk} (assunzione: tutti diversi)
Y = {y1, y2, …, yk}
Pr(x1 uguale a y1) = 1/n
Pr(x1 diverso da y1) = 1-1/n
Pr(Y non comprende x1) = $(1-1/n)^k$

P(almeno [illegible]mento comune in due insiemi di k elementi scelti tra n) = P'(n,k) =

$= 1- ((1- 1/n)^k)^k > 1- ((e^{-1/n})^k)^k = 1-e^{-k^2/n}$.

$P'(n,k) > 0.5$ per $k > 0.83*\sqrt{n}$

$P(n,k) > 0.5$

$1 - e^{-k^2/n} > 0.5$

$- e^{-k^2/n} > - 0.5$

$e^{-k^2/n} < 1/2$

$e^{k^2/n} > 2$

$\ln(2) < k^2/n$

$k > \sqrt{(\ln(2)n)}$

$k > 0.83*\sqrt{n}$

# il problema è mente diverso:

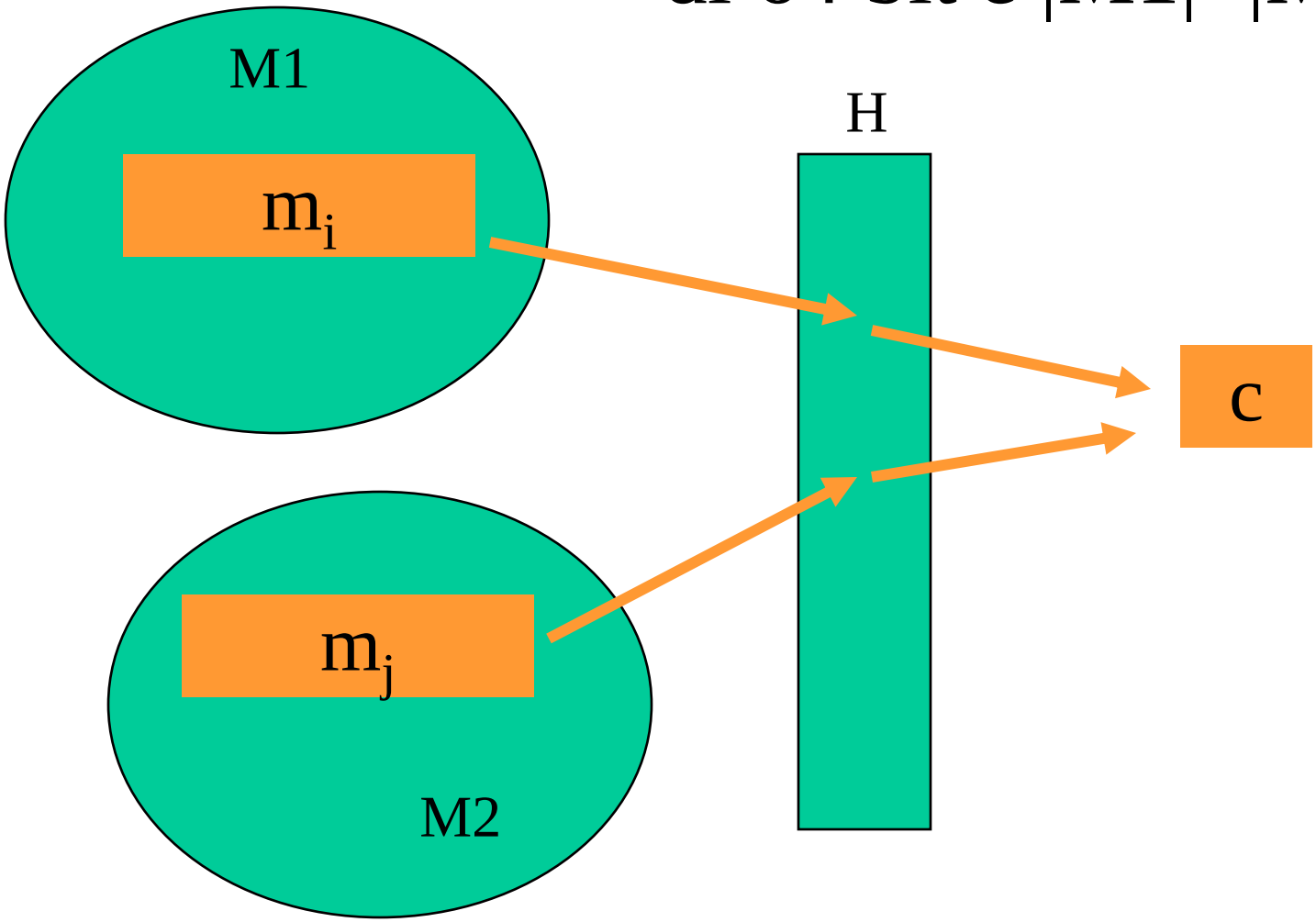
ento comune in due insiemi

i tra n) = P’(n,k) =

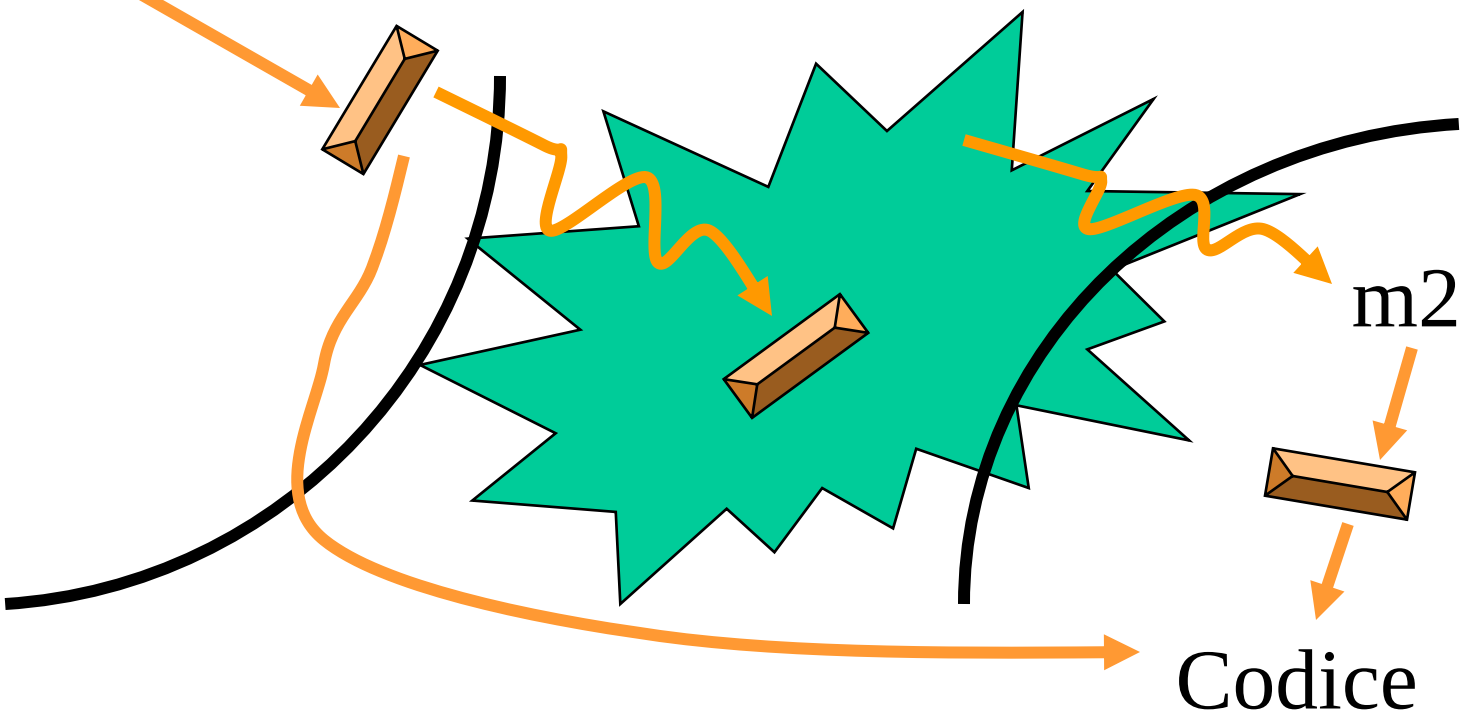
= 1- ((1- 1/n)$^{k}$)$^{k}$ > 1- ((e$^{-1/n}$)$^{k}$)$^{k}$ = 1-e$^{-k^2/n}$.

P’(n,k) > 0.5 per $k > 0.83*\sqrt{n}$

Con probabilità > 0.5, se c di 64 bit e $|M1|=|M2|= 2^{32}$



E’ ora possibile falsificare messaggi:




m1∈M1 = messaggi accettabili per mittente

m2∈M2 = msg equivalenti al msg falso desiderato

# Attacco del compleanno

Eva trova una variante $m_i$ di M

e una variante $m'_j$ di M' tali che

$<m_i, m'_j>$ sia una collisione per H

Eva fa riautenticare ad Alice $m_i$ (ad esempio con firma elettronica)

Eva usa l'autenticazione di $m_i$ per $m'_j$

# Attacco del compleanno

Definizioni di:

Chosen message attack

Known message attack

Ciphertext only attack

# Attacco del compleanno

Serve in questo caso un contesto di

Chosen message attack

Perché Eva “sceglie” $m_i$ e lo fa riautenticare ad Alice (cosa non necessariamente possibile in pratica).

# … quindi

Occorre scegliere funzioni di hash H dove

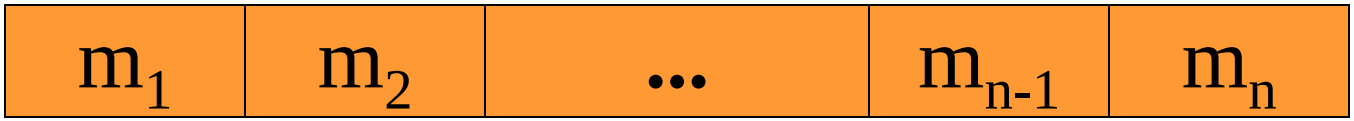
|H(m)| = k, e non è possibile calcolare H per più di $2^{k/2}$ messaggi.

Per questo la lunghezza dei valori di Hash è solitamente 128 bit o più (128 per MD5, 160 per SHA-1).

# Funzioni di hash

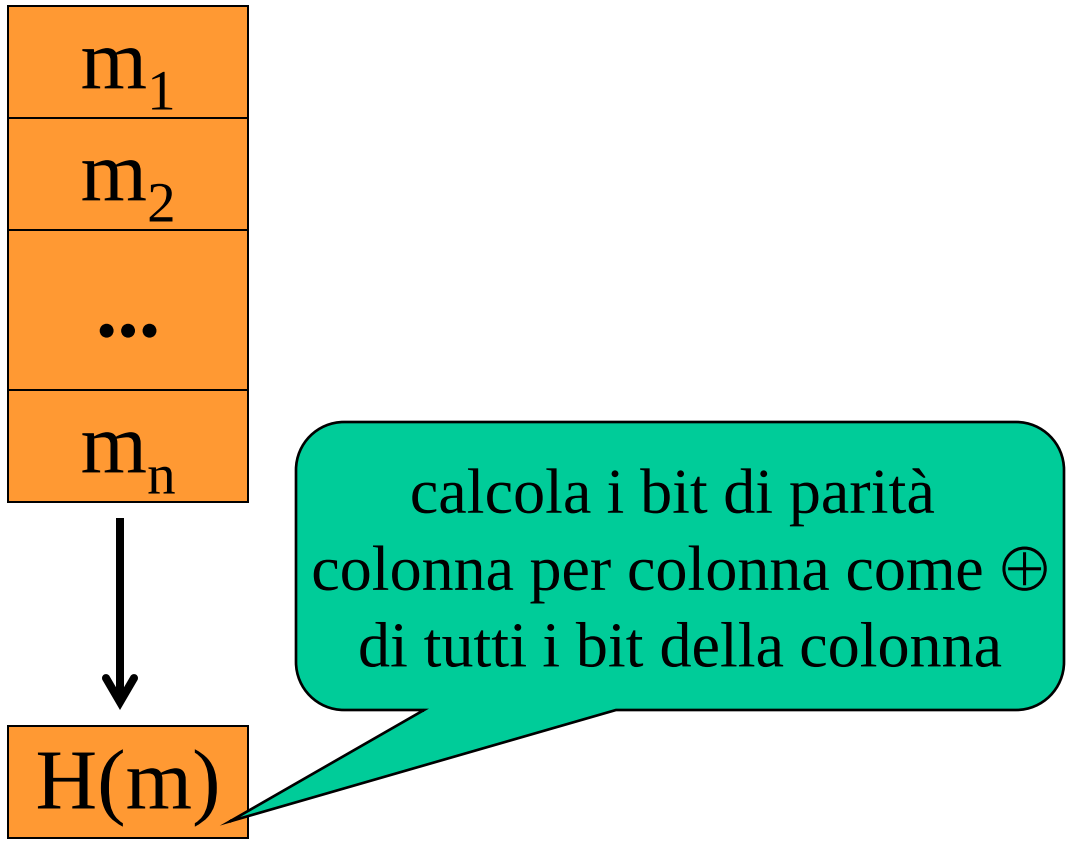
**La lunghezza del digest non basta tuttavia a garantire la sicurezza della funzione di hash.**

# Esempio: H(m) = blocco di parità pari

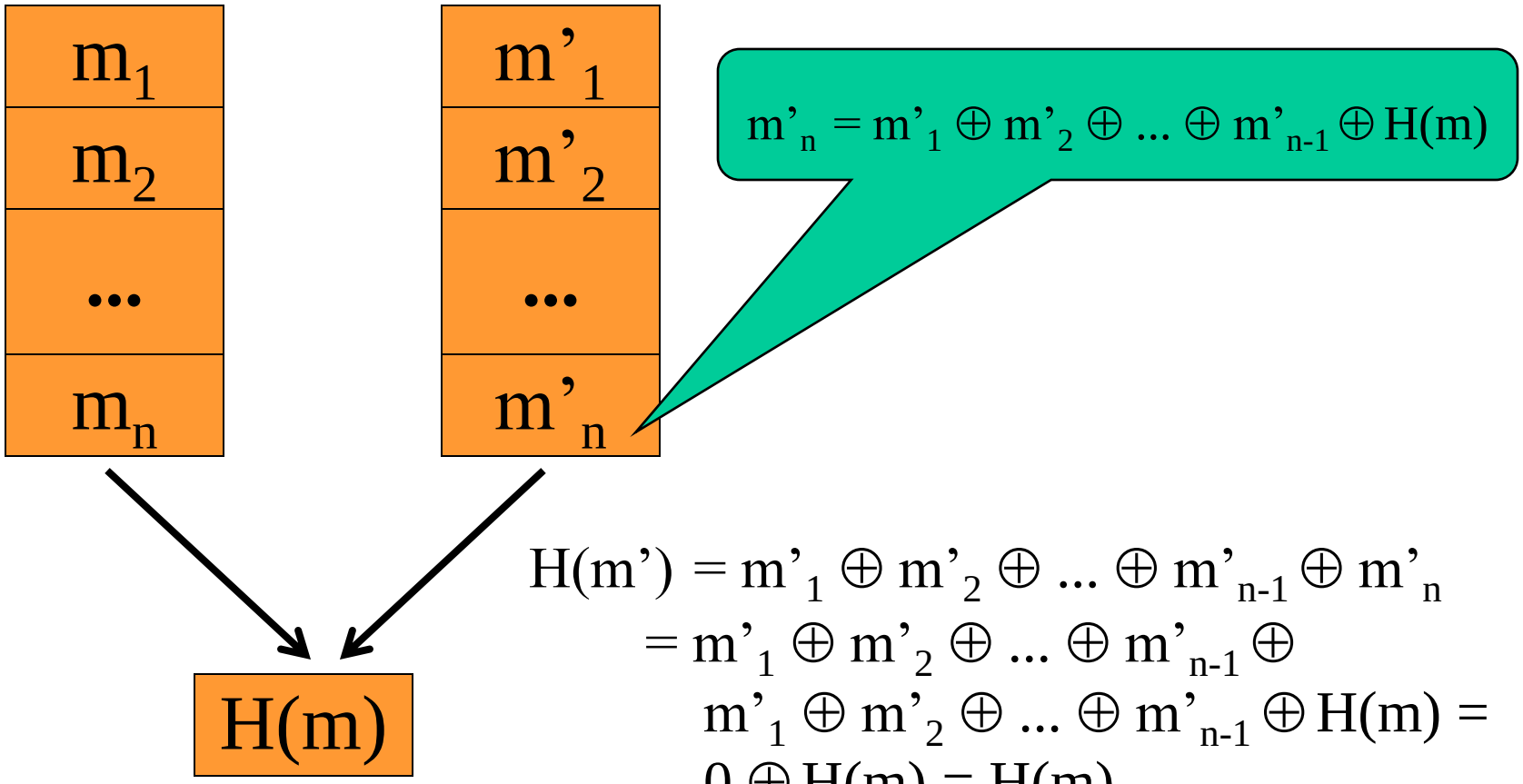
**m =** 


Ogni $m_i$ è lungo k bit con k > 128 bit

# Esempio: H(m) = blocco di parità pari



# Generazione di una collisione



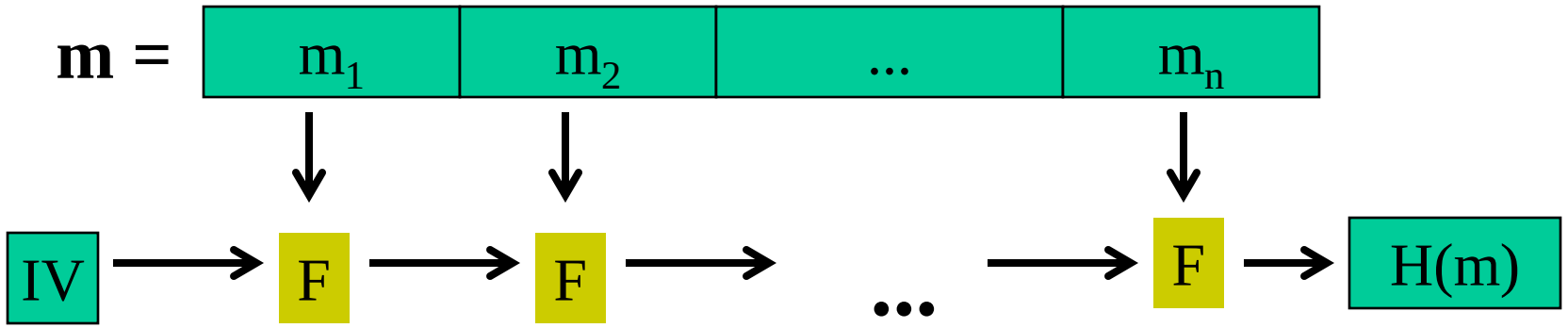

$$H(m') = m'_1 \oplus m'_2 \oplus ... \oplus m'_{n-1} \oplus m'_n$$
$$= m'_1 \oplus m'_2 \oplus ... \oplus m'_{n-1} \oplus$$
$$m'_1 \oplus m'_2 \oplus ... \oplus m'_{n-1} \oplus H(m) =$$
$$0 \oplus H(m) = H(m)$$

# Sicurezza delle funzioni di hash

**Occorrono quindi funzioni di hash che**

- **producano digest sufficientemente lunghi**
- **non permettano semplici metodi di generazione di collisioni**

# Schema generale delle funzioni di hash più usate

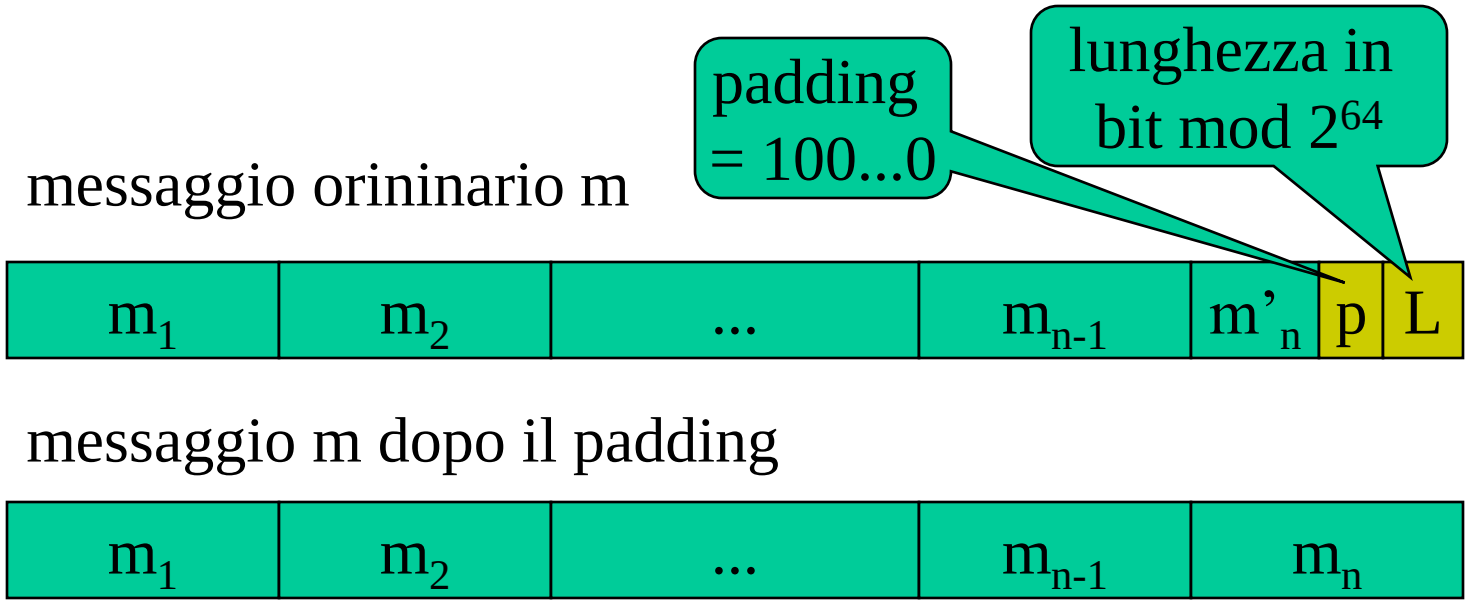



Osservazione [Damgard 1989]: se F è resistente alle collisioni rispetto a messaggi m di lunghezza fissa, allora lo schema complessivo H è resistente alle collisioni rispetto a messaggi di lunghezza arbitraria

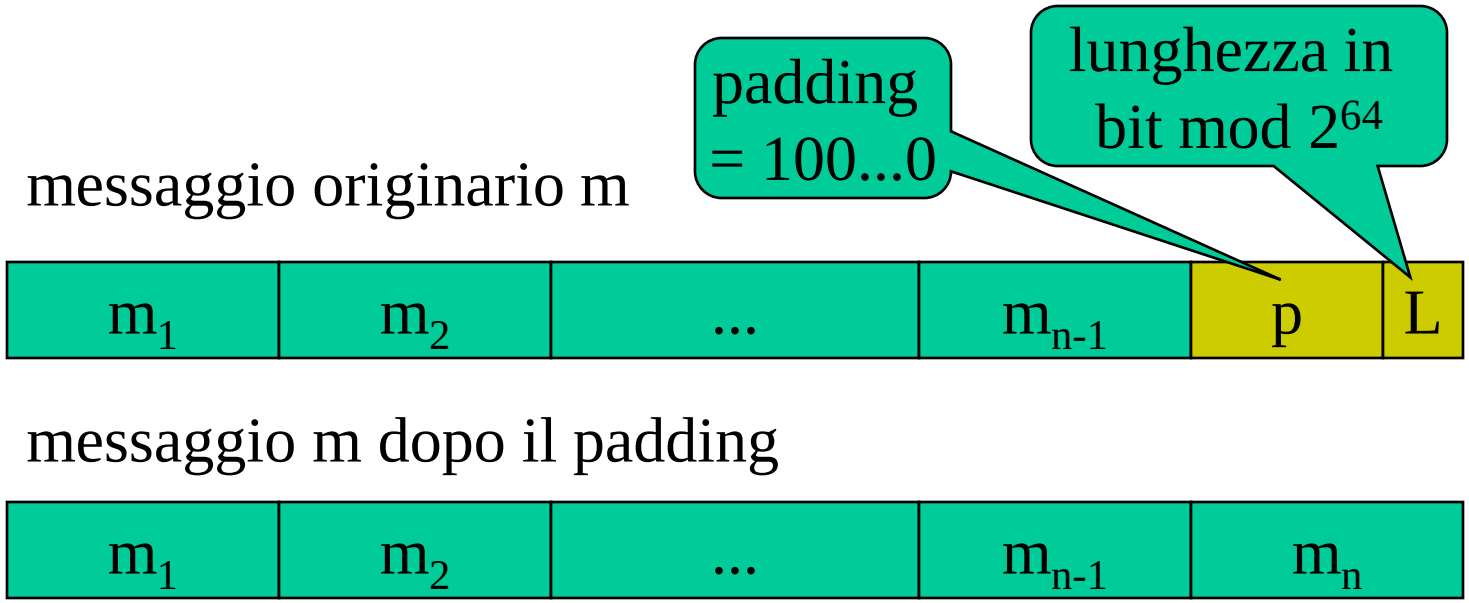
# Funzioni di hash più utilizzate

- **MD5 (Message Digest 5)**
- **SHA-1 (Secure Hash Algorithm versione 1)**

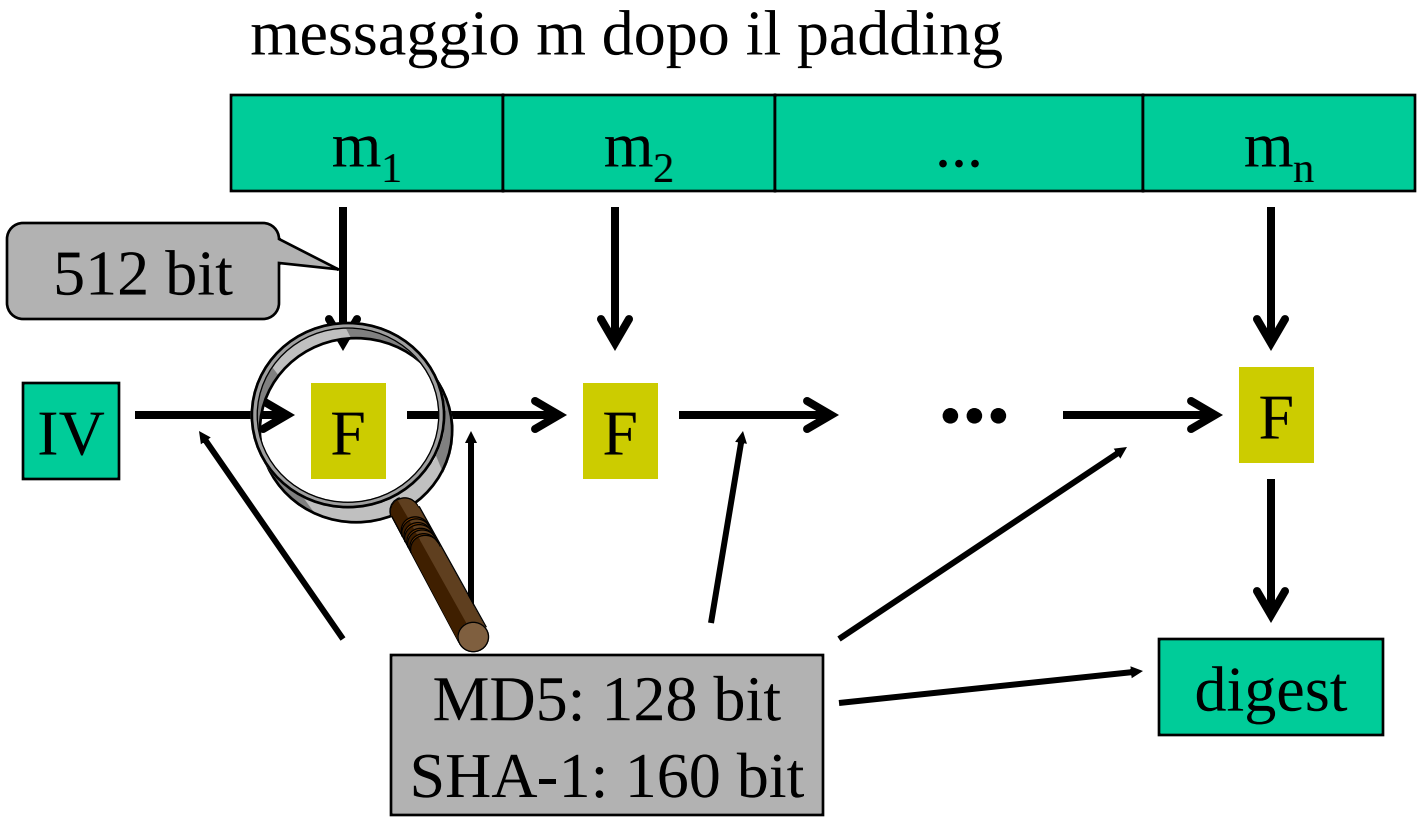
# Schema generale di MD5 e SHA-1



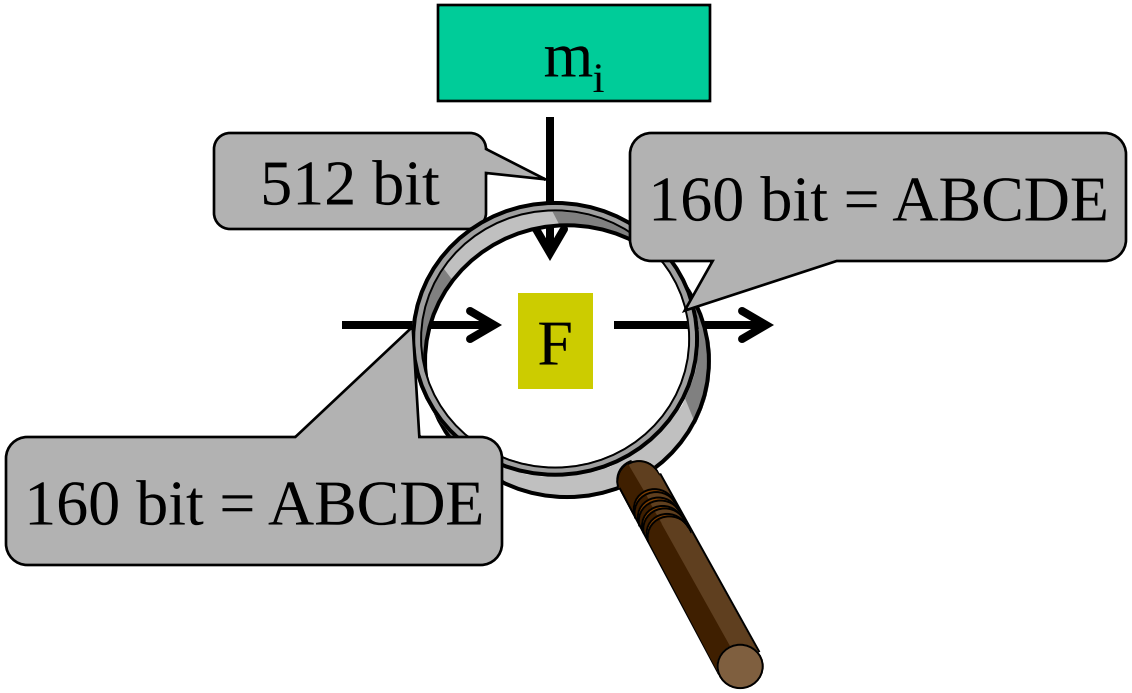
# Schema generale di MD5 e SHA-1



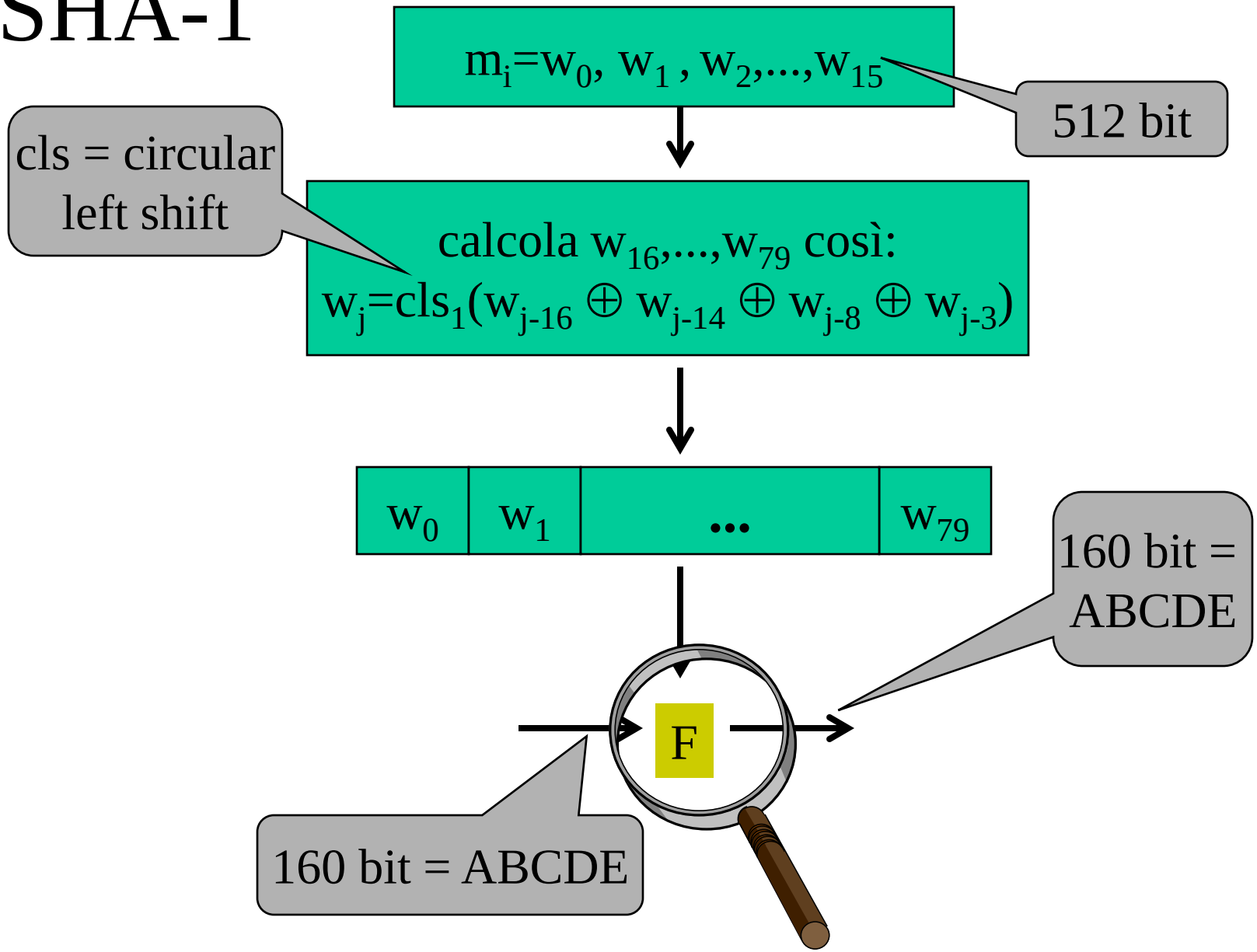
# Schema generale di MD5 e SHA-1



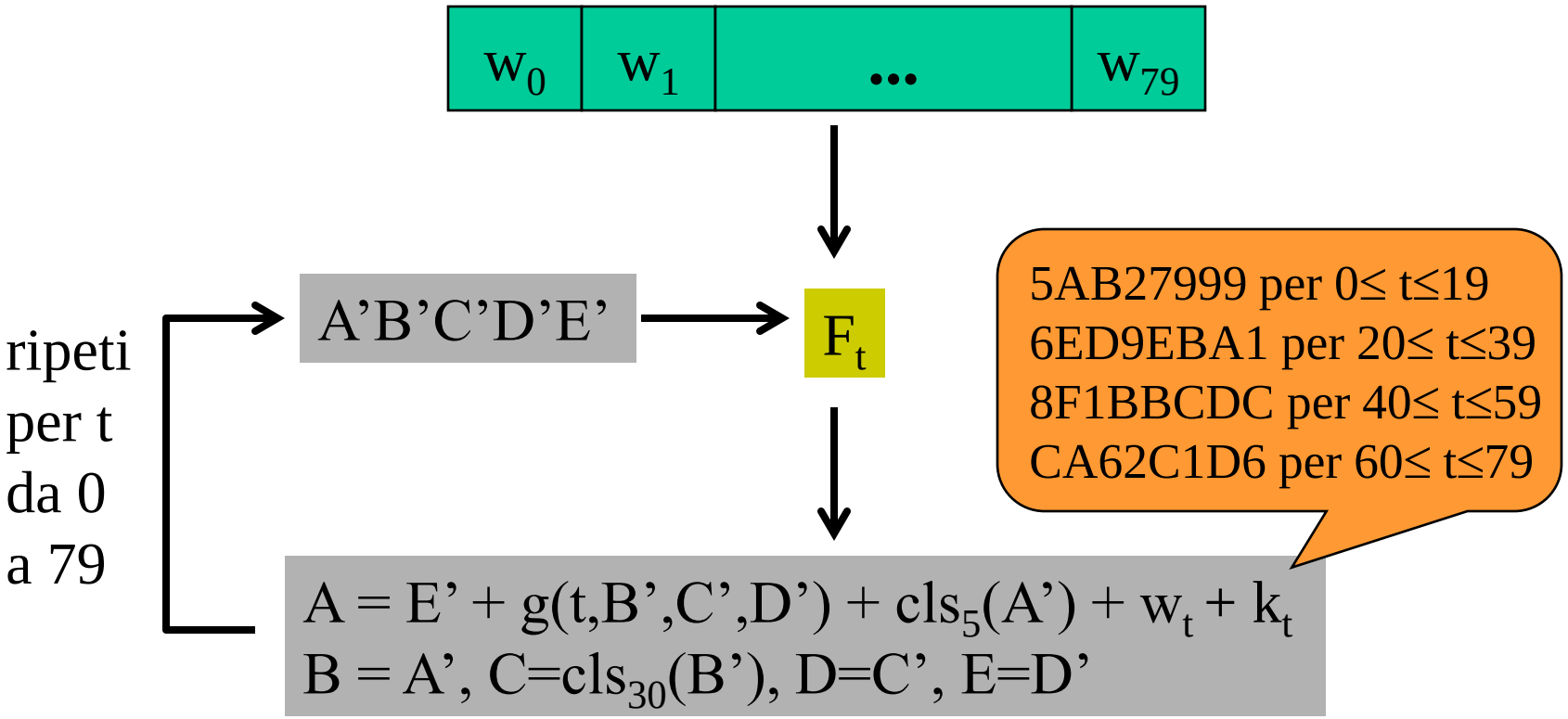
# Funzione di compressione di SHA-1



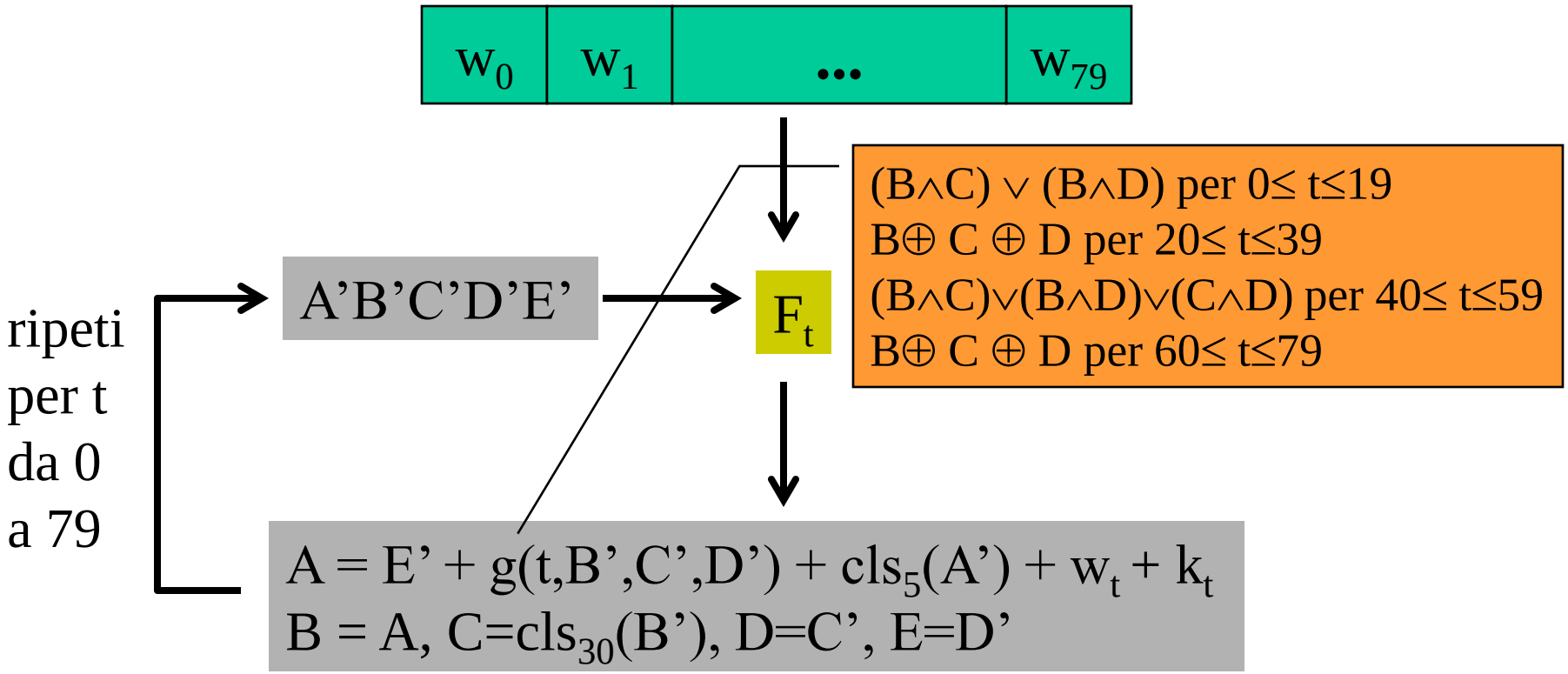
# SHA-1


$m_i=w_0, w_1, w_2,\ldots,w_{15}$
512 bit
cls = circular left shift
calcola $w_{16},\ldots,w_{79}$ così:
$w_j=cls_1(w_{j-16} \oplus w_{j-14} \oplus w_{j-8} \oplus w_{j-3})$
$w_0$
$w_1$
...
$w_{79}$
160 bit = ABCDE
F
160 bit = ABCDE

# Funzione di compressione di SHA-1



# Funzione di compressione di SHA-1



# Caratteristiche di SHA-1 e MD5

- Considerate **non invertibili** o a una via (one-way): dato c, è difficile trovare m tale che H(m) = c
- Probabilmente **resistenti alle collisioni** (collision resistant): è difficile trovare m1 e m2 tali che H(m1) = H(m2)

# Caratteristiche di SHA-1 e MD5

- E’ stato trovato un modo per generare collisioni su un singolo blocco di 512 bit per MD5 [Dobbertin], non così per SHA-1
- SHA-1 più sicuro contro attacchi di tipo ‘forza bruta’ (birthday attacks) perché la lunghezza del digest è maggiore

# Caratteristiche di SHA-1 e MD5

SHA-1 e MD5 utilizzati in

- standard per firma elettronica
- tutti i pacchetti crittografici
- nei più diffusi Browser/MUA e Server Web