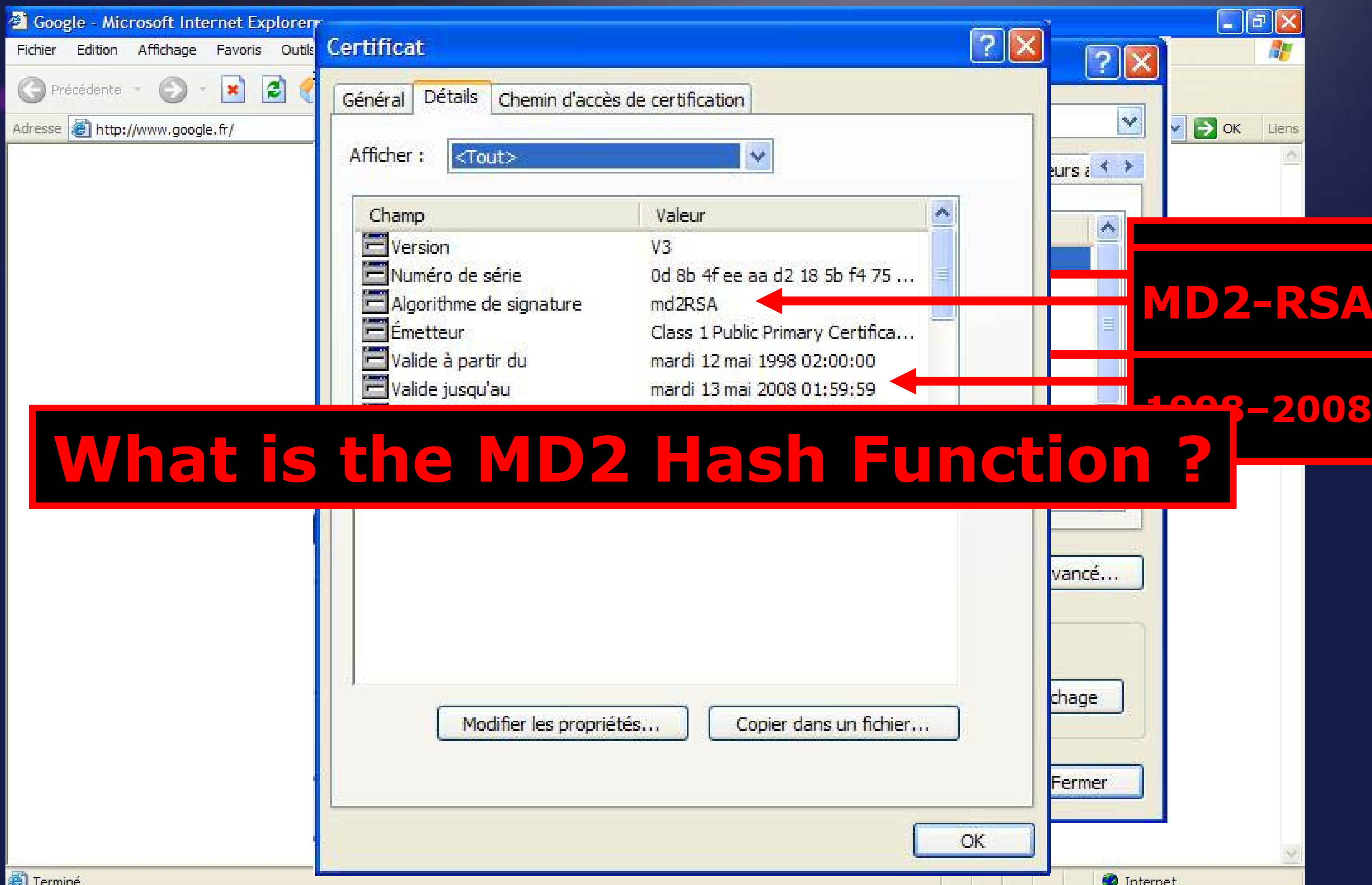
The MD2 Hash Function is not One-Way

Frédéric Muller
D.C.S.S.I. Crypto Lab

A Concrete Situation



What is the MD2 Hash Function ?

Popular Hash Functions

- The SHA family (developed by NIST)
  - SHA-0 (collision found in August 2004)
  - SHA-1
  - SHA-256 and sisters
- The MD Family (developed by RSA Labs)
  - MD2
  - MD4 (collision found in 1996)
  - MD5 (collision found in 2004)
- Other algorithms
  - RIPEMD
  - HAVAL

The MD2 Hash Function

- It was designed by Ron Rivest in 1989 (published in a 1992 RFC)
- Non-classical construction (early design)
- Part of PKCS #1 v1.5 and 2.1 standards
- Few cryptanalysis results :
  - Collision on a simplified version (Rogier-Chauvaud, 1995)

Results in this paper

Important weaknesses of MD2 :

- The compression function can be inverted with complexity 2^{73} basic operations (meet-in-the-middle attack)
- Consequence = Preimage and Second preimage attacks cost 2^{104}

$\Rightarrow$ MD2 is not a secure One-Way Hash

Hash Functions

- Input = a message of arbitrary length
- Output = a hash of fixed size (128 bits for MD2)

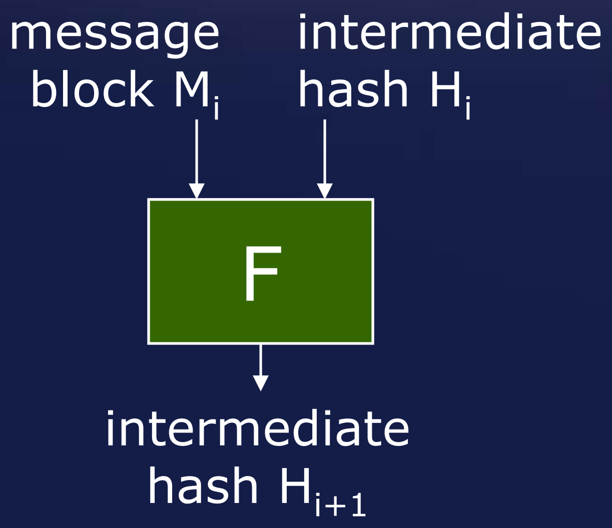
$$H : \{0,1\}^* \longrightarrow \{0,1\}^{128}$$

Security of Hash Functions

- Collision resistance
  - It should be difficult to find M and M′ such that H(M) = H(M′)
- Second preimage resistance
  - For a given M, it should be difficult to find M′ such that H(M) = H(M′)
- Preimage resistance
  - For a given h, it should be difficult to find M such that H(M) = h

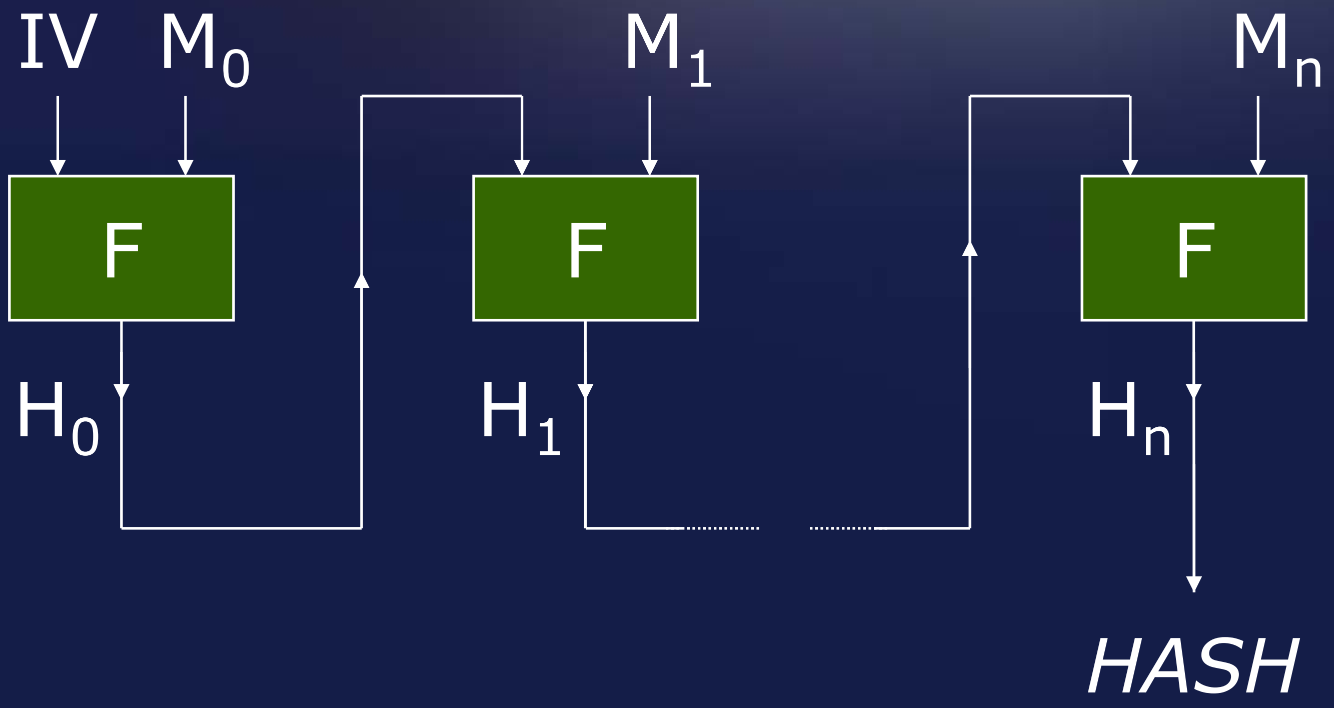
Compression Function

- The basic tool is a compression function F



- Message blocks have length 128 bits for MD2.

Iterated Hash Functions



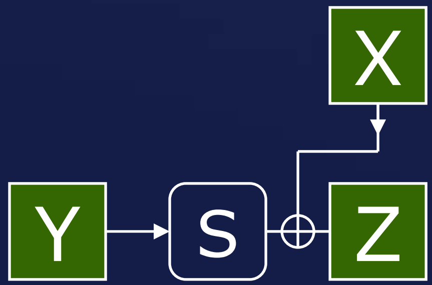
Particularities of MD2

- not Merkle-Damgaard
  - → Last message block = non-linear checksum
- not Davies-Meyer
  - → Dedicated compression function
- All operations are byte-oriented

A basic tool

The basic function is

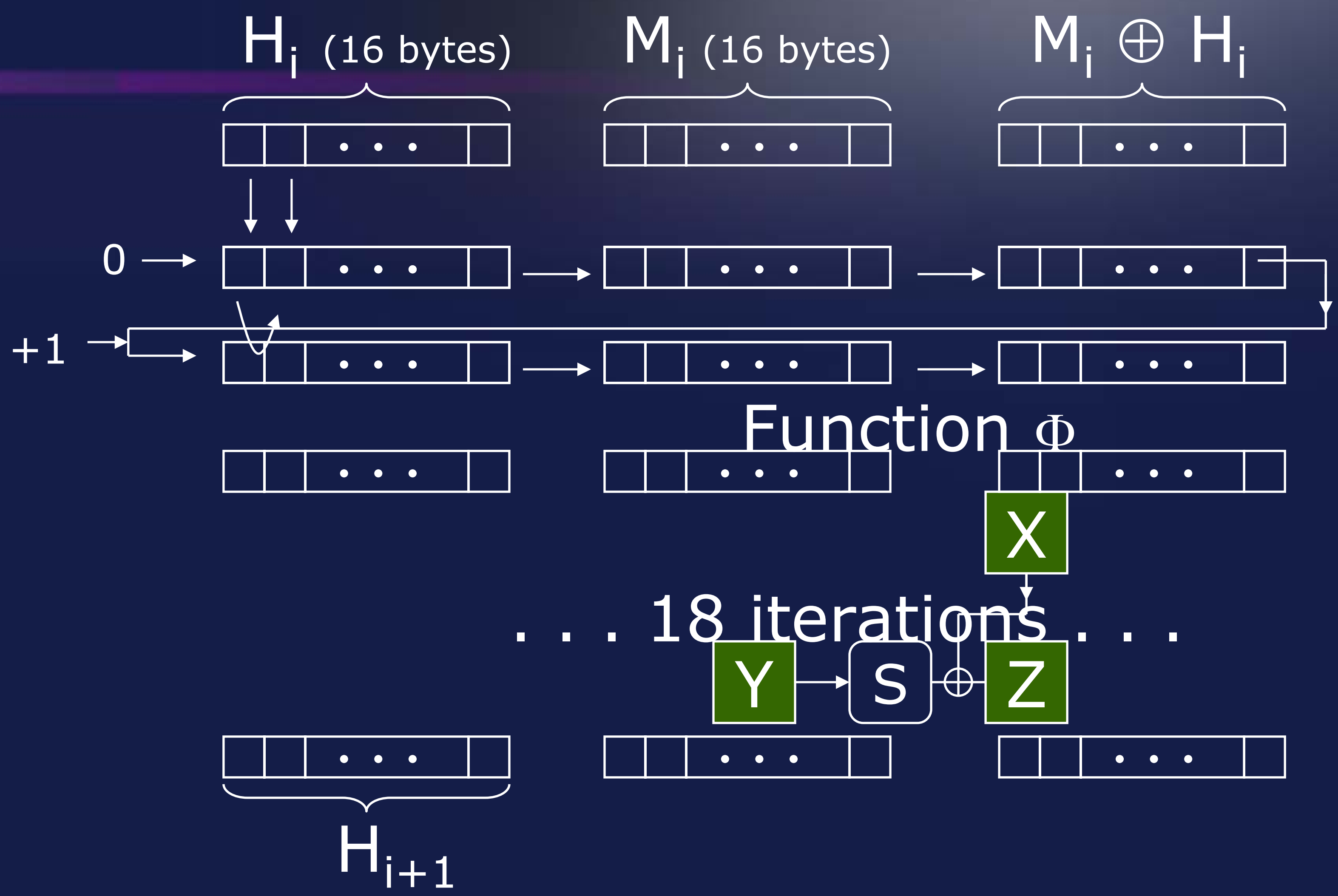
$$\Phi(X,Y) = Z = X \oplus S(Y)$$



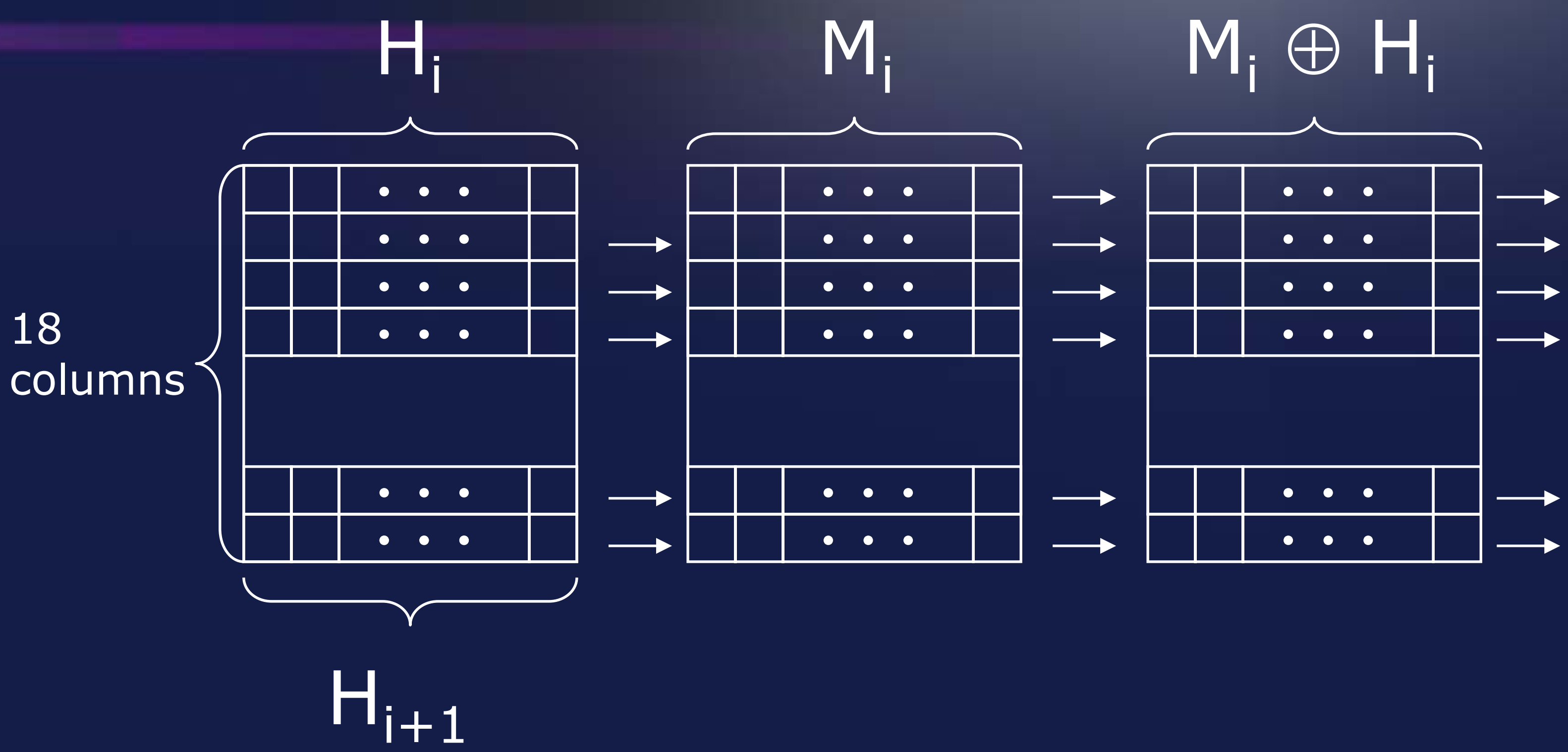
where S is a 8→8 S-box

Φ is invertible when one input is known

MD2 compression function



Representation



Intermediate values are stored in 3 matrices

Attacks against F

$$H_{i+1} = F(H_i, M_i)$$

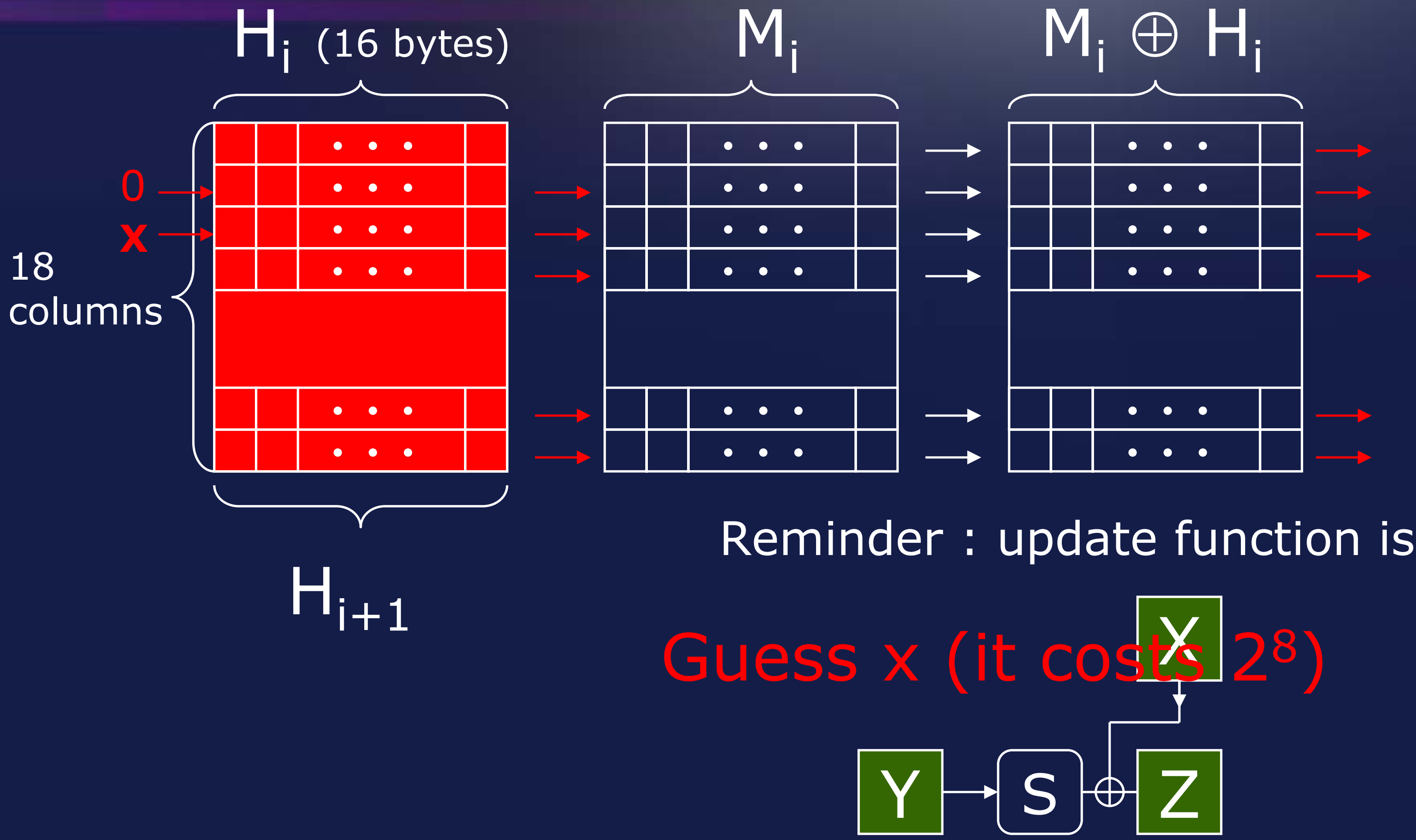
2 "preimage" attacks against F :

- H_i and H_{i+1} are given, find M_i
  Complexity 2^{95}
- H_{i+1} is given, find M_i and H_i
  Complexity 2^{73}

General Ideas of these Attacks

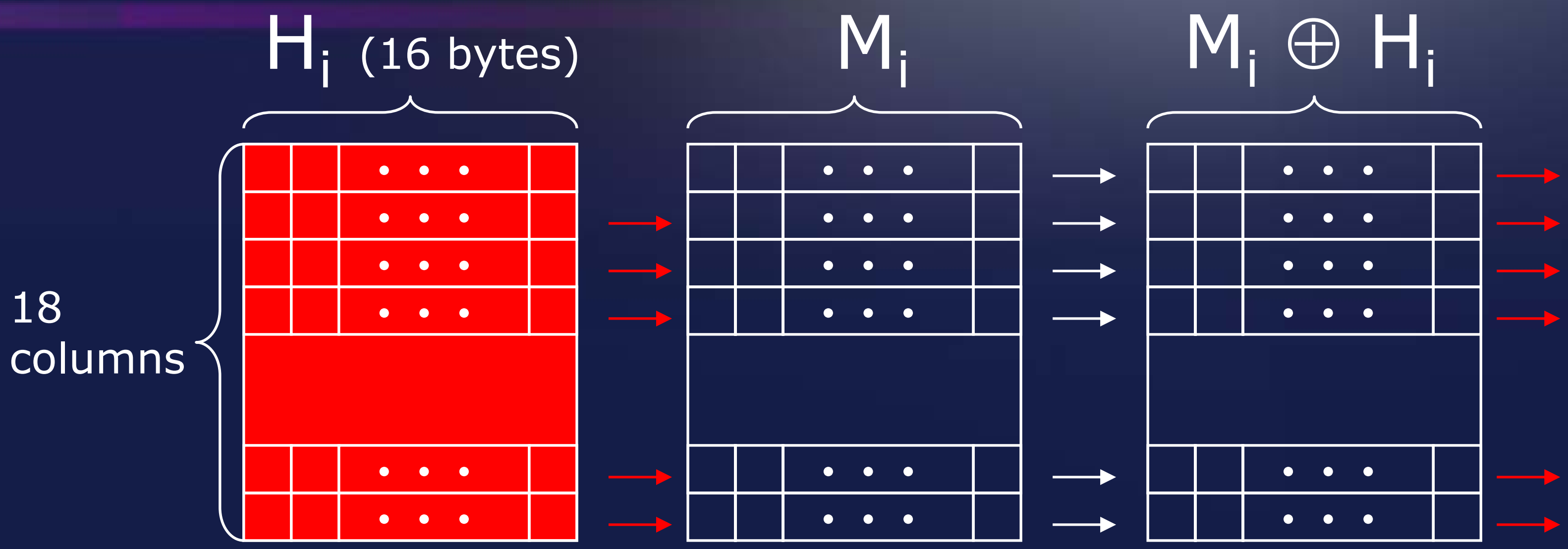
1. Determine portions of the state from known values (like H_{i+1})
   $\Rightarrow$ indeed Φ is "invertible"
2. Guess separately the two halves of the unknown.
3. "meet-in-the-middle" : find a match ($\approx$ solution)

When H_i and H_{i+1} are given



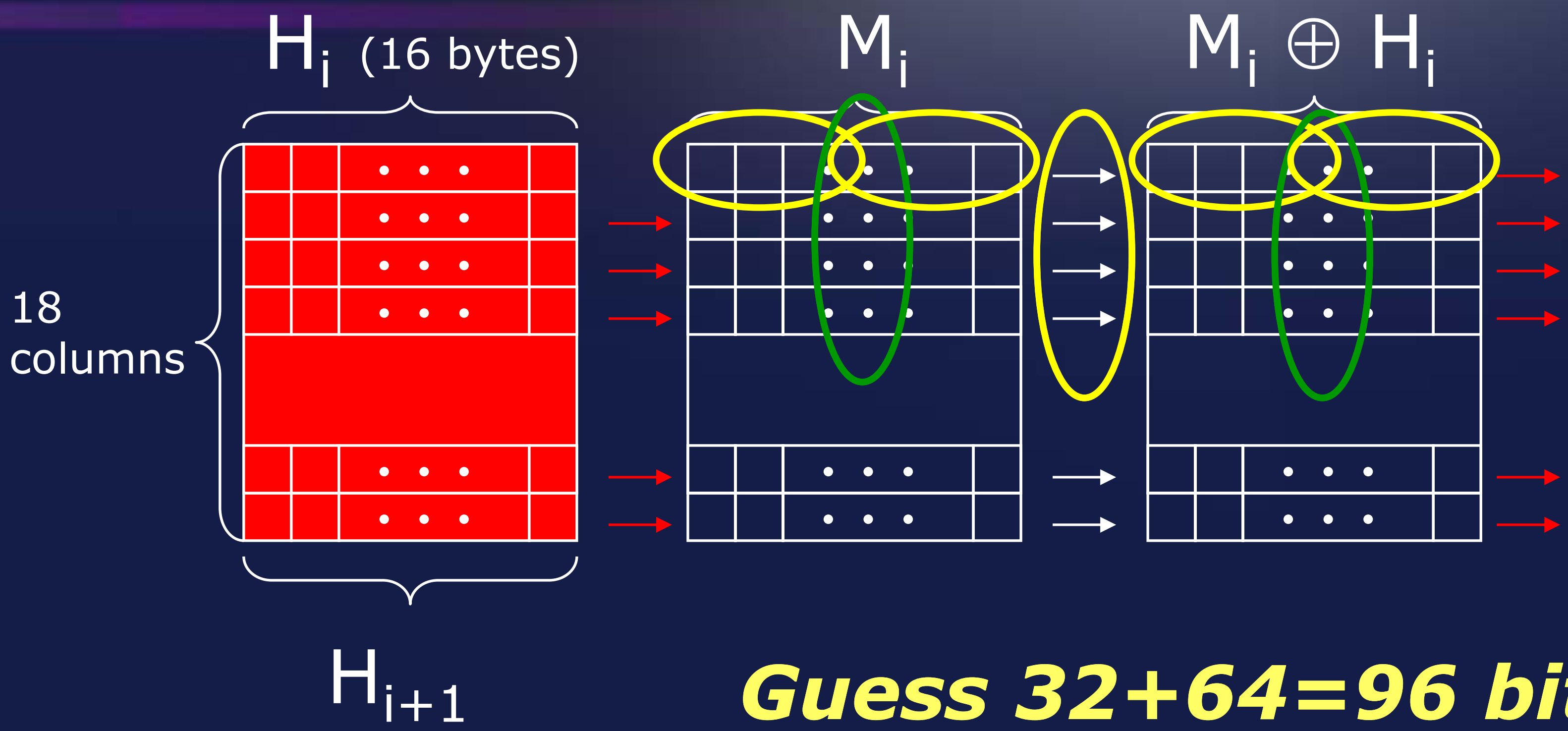
Guess x (it costs 2^8)

General Idea



- *Guess the left half of M_i*
- *Guess the right half of M_i*
- *Match intermediate values «in the middle»*

“Meet-in-the-middle” attack



Guess 32+64=96 bits

Determine 64 bits

Summary

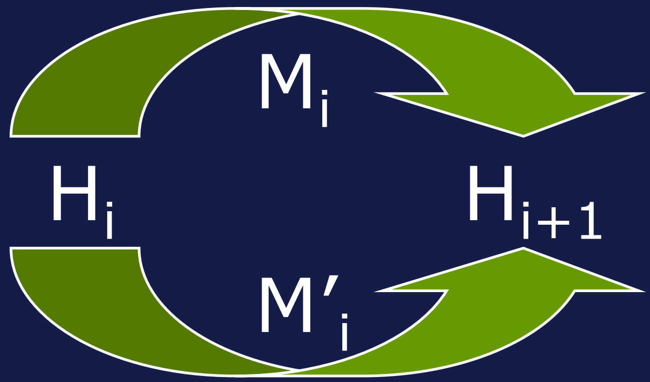
- This attack costs roughly $2^{96} \times 2^{8} = 2^{104}$
- Works when H_i and H_{i+1} are given, it retrieves ALL acceptable solutions M_i
- When only H_{i+1} is given, a similar attack finds an acceptable (H_i, M_i) costs 2^{73}

Application to the whole hash

- Merkle-Damgaard : attacks against F turn into attacks against the whole hash
- Here : last block of message must match the non-linear checksum
- Idea : multi-collisions for hash functions (Joux-04)

Chaining Attack

- Goal = find a preimage of some target x
- Pick a sequence of intermediate hashes $H_0 \ldots H_{128}$ such that
  - H_0 = IV of MD2 = 0
  - $H_{128} = x$
  - Two possible message blocks M_i and M'_i at each step



Chaining Attack

- Apply only 128 times the previous attack against F
- All messages map to x
  $\Rightarrow$ we get "almost for free" 2^{128} preimages instead of just 1

Chaining Attack

- 2^{128} different preimages of x
- One should verify the checksum constraint
- Costs 2^{64} to identify
- Overall Complexity

  = 128 attacks against F

  $\approx 2^{104}$

Conclusion

- Preimage and second preimage Attacks for MD2 faster then 2^{128} (not practical yet)
- MD2 is not a secure one-way hash function
- General results (Kelsey/Schneier) do not apply well because MD2 is not Merkle-Damgaard