Second Preimage Attacks on Dithered Hash Functions

Elena Andreeva[1] Charles Bouillaguet[2]
Pierre-Alain Fouque[2] Jonathan J. Hoch[3] John Kelsey[4]
Adi Shamir[2,3] Sebastien Zimmer[2]

[1]K.U. Leuven, ESAT/COSIC, Leuven-Heverlee, Belgium

[2]École Normale Supérieure, Paris, France

[3]Weizmann Institute of Science, Rehovot, Israel

[4]NIST, Gaithersburg, MD, USA

EUROCRYPT 2008

Hash Functions Cryptanalysis

$$H : \{0,1\}^* \mapsto \{0,1\}^n$$

Should behave “like a random oracle”.

Hash Functions Cryptanalysis

$$H : \{0,1\}^* \mapsto \{0,1\}^n$$

Should behave "like a random oracle".

Collision attack Find $M_1 \neq M_2$ s.t. $H(M_1) = H(M_2)$.
Ideal security: $2^{n/2}$.

Second-preimage attack Given M_1, find $M_2 \neq M_1$ s.t. $H(M_1) = H(M_2)$.
Ideal security: 2^n.

Preimage attack Given y, find M s.t. $H(M) = y$.
Ideal security: 2^n.

Hash Functions Cryptanalysis

$$H : \{0,1\}^* \mapsto \{0,1\}^n$$

Should behave “like a random oracle”.

Collision attack Find $M_1 \neq M_2$ s.t. $H(M_1) = H(M_2)$.
Ideal security: $2^{n/2}$.

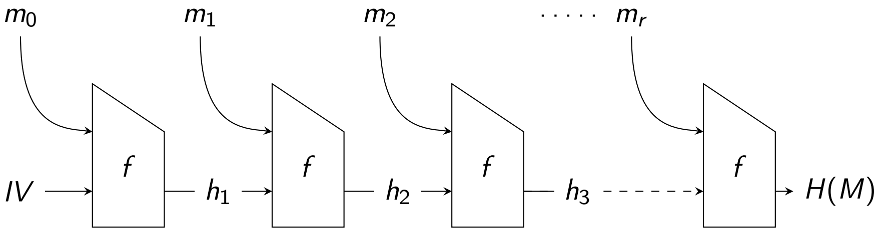
Second-preimage attack Given M_1, find $M_2 \neq M_1$ s.t. $H(M_1) = H(M_2)$.
Ideal security: 2^n.

Preimage attack Given y, find M s.t. $H(M) = y$.
Ideal security: 2^n.

The Merkle-Damgård Mode of Operation

Most hash functions are iterated hash functions :

- Split M into m-bit blocks : $M = m_0, m_1, \ldots, m_r$
- Pad the last block (include binary encoding of $|M|$)
- Iterate a compression function $f : \{0,1\}^{n+m} \to \{0,1\}^n$



Generic Attacks

A full hash function is made of

- A compression function
- A mode of operation (i.e., a way of using it)

In this talk

Attacks against the mode of operation

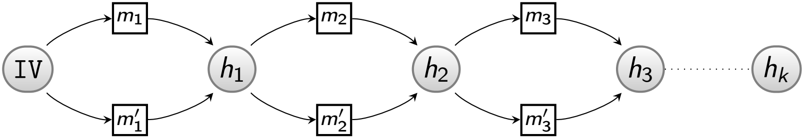
- Works for all f : generic attacks
- Model f as a Random Oracle
- Collisions on f cost $2^{n/2}$

Joux's Multicollision [CRYPTO'04]

Towards the First Generic Second Preimage Attack

For the cost of k collisions, we can build a 2^k-multicollision

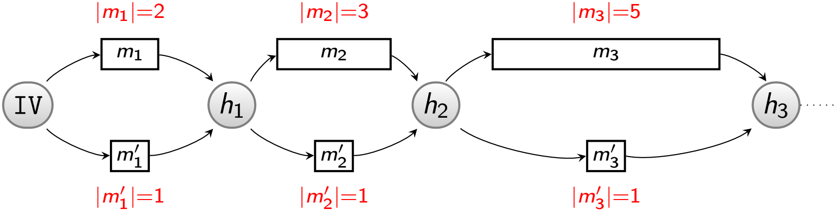
- At each step, find a colliding block pair starting from the last chaining value
- 2^k paths between IV and h_k



Works because of the iterated structure of H !

Kelsey & Schneier Second Preimage Attack [EUROCRYPT'05]

At step i, find a collision between a 1-block message and a $(2^i + 1)$-block message



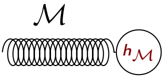
- Messages of sizes $[k+1; 2^{k+1} - 2]$ that hash to h_k

$\Rightarrow$ expandable message

How to use this ?

Kelsey & Schneier Second Preimage Attack (Cont'd)

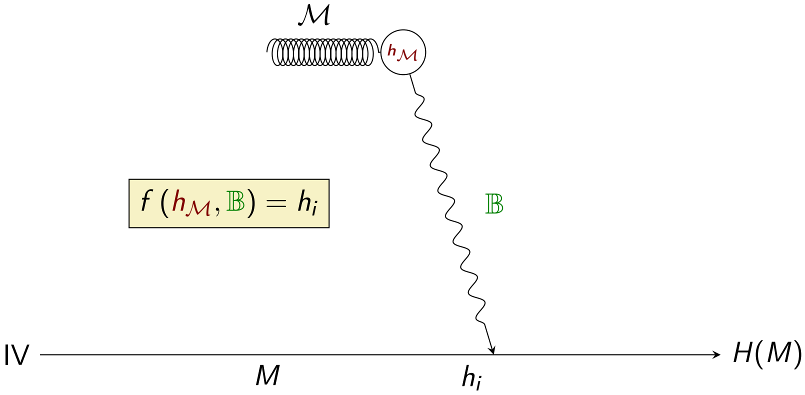
1. Generate an Expandable Message $\mathcal{M}$ that hashes to $h_{\mathcal{M}}$



$$\mathrm{IV} \xrightarrow[M]{\hspace{6em}} H(M)$$

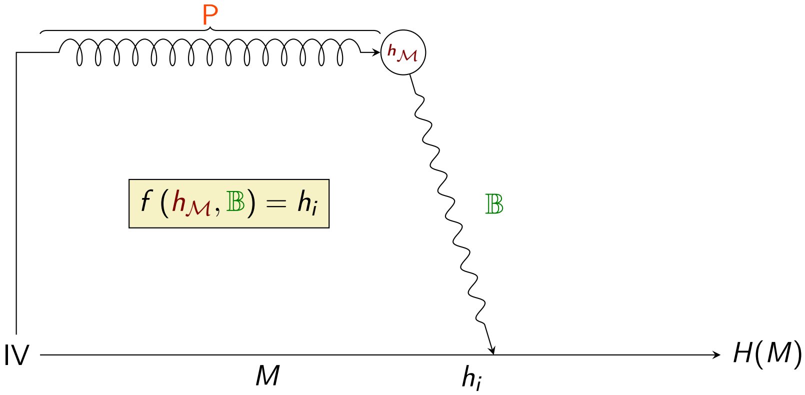
Kelsey & Schneier Second Preimage Attack (Cont'd)

1. Generate an Expandable Message $\mathcal{M}$ that hashes to $h_{\mathcal{M}}$
2. Find a message block $\mathbb{B}$ "connecting" $h_{\mathcal{M}}$ to M



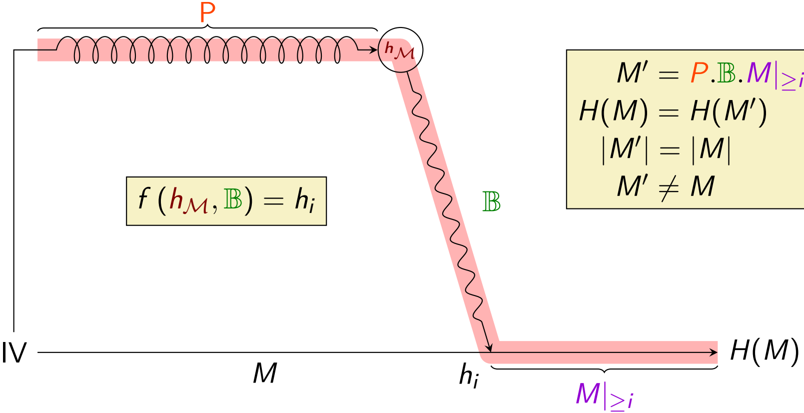
Kelsey & Schneier Second Preimage Attack (Cont'd)

1. Generate an Expandable Message $\mathcal{M}$ that hashes to $h_{\mathcal{M}}$
2. Find a message block $\mathbb{B}$ "connecting" $h_{\mathcal{M}}$ to M
3. Using $\mathcal{M}$, build P of length $i-1$ that hashes to $h_{\mathcal{M}}$



Kelsey & Schneier Second Preimage Attack (Cont'd)

1. Generate an Expandable Message $\mathcal{M}$ that hashes to $h_{\mathcal{M}}$
2. Find a message block $\mathbb{B}$ "connecting" $h_{\mathcal{M}}$ to M
3. Using $\mathcal{M}$, build P of length $i-1$ that hashes to $h_{\mathcal{M}}$
4. Assemble all pieces to form a second preimage M'



$$\begin{aligned} M' &= P.\mathbb{B}.M|_{\geq i} \\ H(M) &= H(M') \\ |M'| &= |M| \\ M' &\neq M \end{aligned}$$

Kelsey & Schneier Second Preimage Attack (end)

Cost of the attack:

- Build Expandable Message $\mathcal{M}$
  - k collisions
  - $2^k \geq |M|$
  - Cost: $k \cdot 2^{n/2}$
- "Connect" $h_{\mathcal{M}}$ to target message (*i.e.*, find $\mathbb{B}$).
  - Cost : $2^n/|M|$.

$\Longrightarrow$ If $|M| = 2^k$, total cost : $k \cdot 2^{n/2} + 2^{n-k}$

- SHA-1 ($k = 55, n = 160$), total cost : 2^{106}

Kelsey & Schneier Second Preimage Attack (end)

Cost of the attack:

- Build Expandable Message $\mathcal{M}$
  - k collisions
  - $2^k \geq |M|$
  - Cost: $k \cdot 2^{n/2}$
- "Connect" $h_{\mathcal{M}}$ to target message (*i.e.*, find $\mathbb{B}$).
  - Cost : $2^n / |M|$.

$\Longrightarrow$ If $|M| = 2^k$, total cost : $k \cdot 2^{n/2} + 2^{n-k}$

- SHA-1 ($k = 55, n = 160$), total cost : 2^{106}

Conclusion

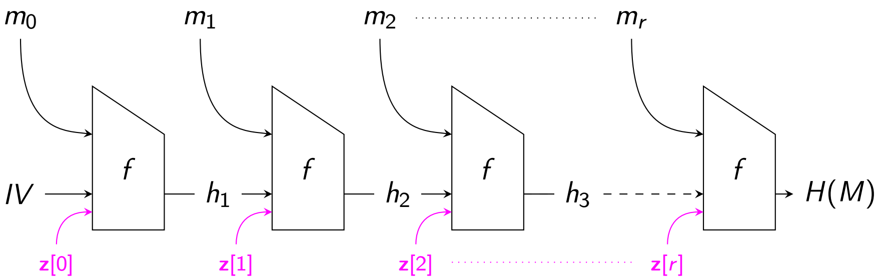
There is a problem with the Merkle-Damgård mode of operation

Dithering

Several new modes of operation recently suggested to replace MD.

- Some prevent the 2nd Preimage attack with dithering.
  - Perturb the hash process
  - new input from a fixed dithering sequence $\mathbf{z}$.
- HAIFA : dithering with a 64-bit counter
- Rivest : dithering with 2-bit symbols

  (Proposed at the 1st NIST Hash Workshop)



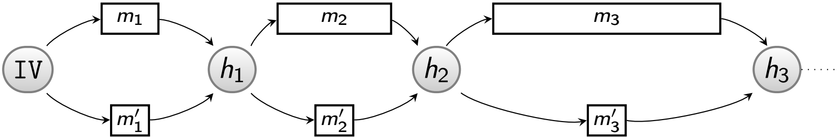
Rivest's Dithering Proposal

Description

Dithering with a repetition-free sequence on 4 letters :

$$\mathbf{z} = abcacdcbcd\,cadcdbdabacabadbabcbdbcba\ldots$$

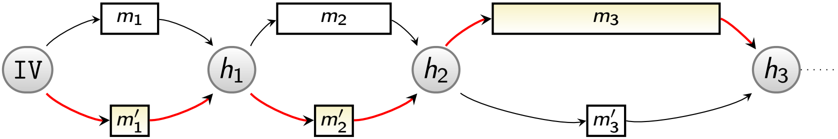
- no square in sequence
  - square : bana.na
- Perturbs construction of the Expandable Message



Rivest's Dithering Proposal
Effectiveness

$\mathbf{z} = abcac\,dcbcd\,cadcdbdabacabadbabcbdbcba\ldots$

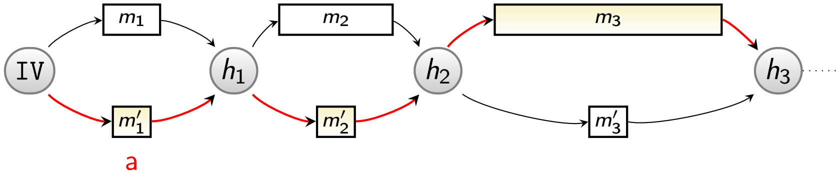
- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m'_2.m_3$
  - e.g. $\ell = 8$. $P = m'_1.m'_2.m_3$



Rivest's Dithering Proposal
Effectiveness

$\mathbf{z} = abcacdcbcd\,cadcdbdabacabadbabcbdbcba\ldots$

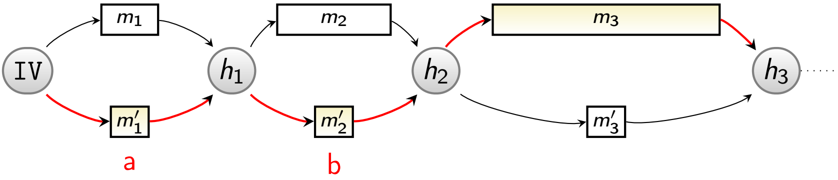
- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$



Rivest's Dithering Proposal
Effectiveness

$\mathbf{z} = abcacdcbcdcadcdbdabacabadbabcbdbcba\ldots$

- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$

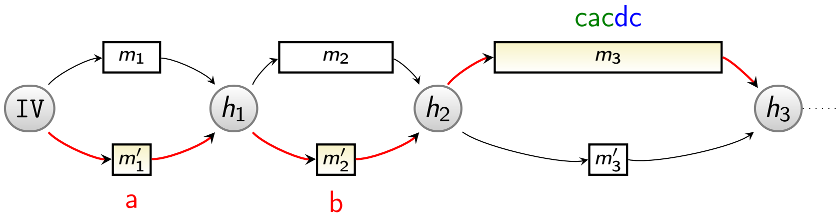


Rivest's Dithering Proposal

Effectiveness

$\mathbf{z} = abcacdcbcd\,cadcdbdabacabadbabcbdbcba\ldots$

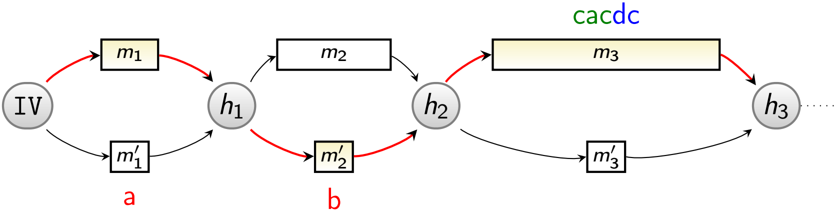
- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$



Rivest's Dithering Proposal
Effectiveness

$\mathbf{z} = abcacdcbcd\,cadcdbdabacabadbabcbdbcba\ldots$

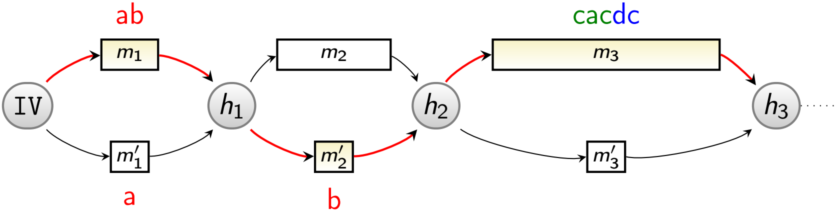
- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$



Rivest's Dithering Proposal
Effectiveness

$\mathbf{z} = abcacdcbcdcadcdbdabacabadbabcbdbcba\ldots$

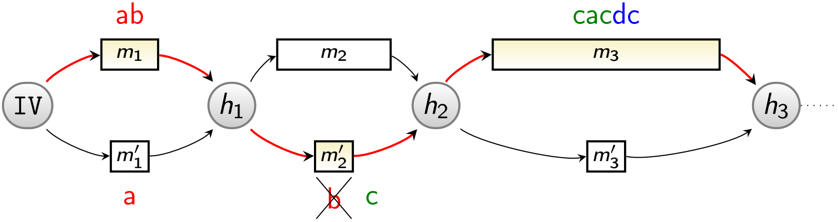
- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$



Rivest's Dithering Proposal
Effectiveness

$\mathbf{z} = abcacdcbcd\,cadcdbdabacabadbabcbdbcba\ldots$

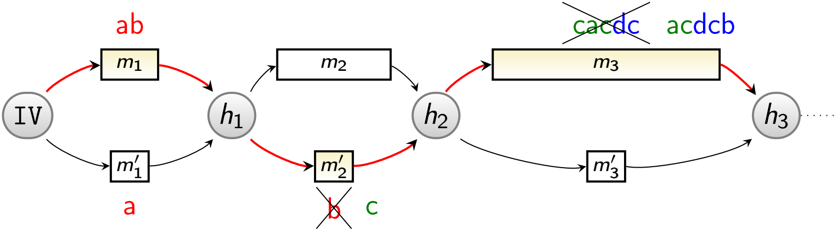
- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$



Rivest's Dithering Proposal
Effectiveness

$z = abcacdcbcdcadcdbdabacabadbabcbdbcba\ldots$

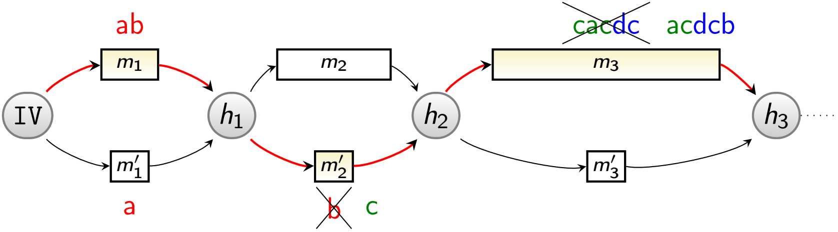
- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$



Rivest's Dithering Proposal
Effectiveness

$\mathbf{z} = abcacdcbcd\,cadcdbdabacabadbabcbdbcba\ldots$

- Need to choose/fix dithering symbols when building $\mathcal{M}$
- How ? Need to match the actual sequence...
  - e.g. $\ell = 7$. $P = m_1.m_2'.m_3$
  - e.g. $\ell = 8$. $P = m_1'.m_2'.m_3$

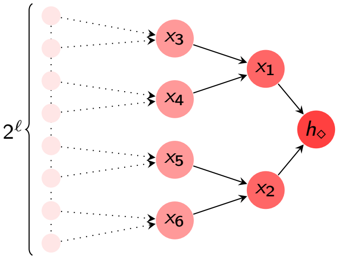


Conclusion

Kelsey and Schneier's attack does not work with dithering

The “Diamond” Structure

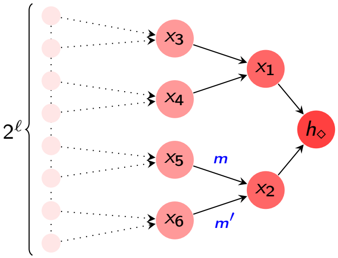
The new attack relies on the diamond structure from the herding attack of Kelsey and Kohno [EUROCRYPT'06].



- Complete binary tree of height ℓ
- Node $\simeq$ chaining values
- Edges $\simeq$ message blocks

The "Diamond" Structure

The new attack relies on the diamond structure from the herding attack of Kelsey and Kohno [EUROCRYPT'06].

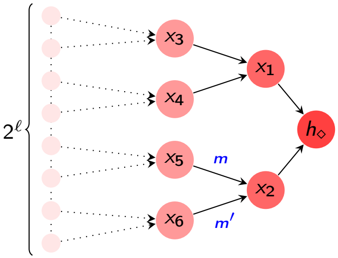


- Complete binary tree of height ℓ
- Node $\simeq$ chaining values
- Edges $\simeq$ message blocks
- Collision tree
- Maps 2^ℓ chaining values to $h_\diamond$ (paths of ℓ blocks in the tree)

$$\boxed{f(x_5, m) = f(x_6, m') = x_2}$$

The "Diamond" Structure

The new attack relies on the diamond structure from the herding attack of Kelsey and Kohno [EUROCRYPT'06].

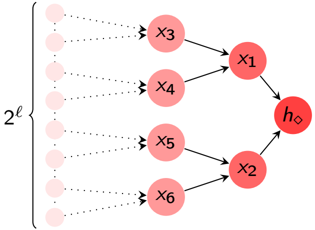


$$\boxed{f(x_5, m) = f(x_6, m') = x_2}$$

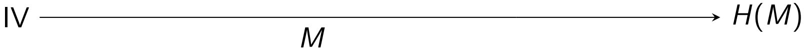
- Complete binary tree of height ℓ
- Node $\simeq$ chaining values
- Edges $\simeq$ message blocks
- Collision tree
- Maps 2^ℓ chaining values to $h_\diamond$ (paths of ℓ blocks in the tree)
- Built in time $2^{n/2+\ell/2+2}$

Putting the "Diamond" at Work

Replaying Kelsey and Schneier's attack, but with a diamond

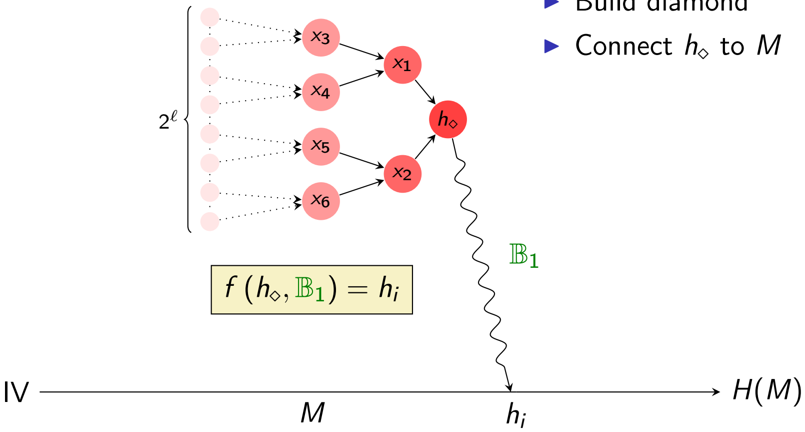


- Build diamond



Putting the "Diamond" at Work

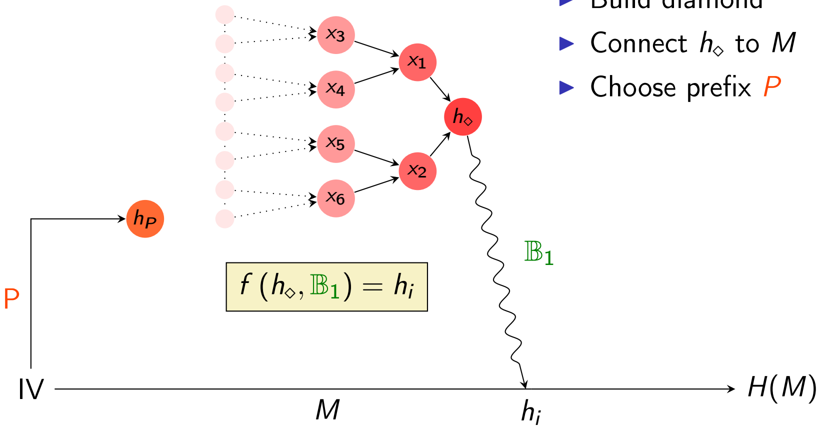
Replaying Kelsey and Schneier's attack, but with a diamond



- Build diamond
- Connect $h_\diamond$ to M

Putting the "Diamond" at Work

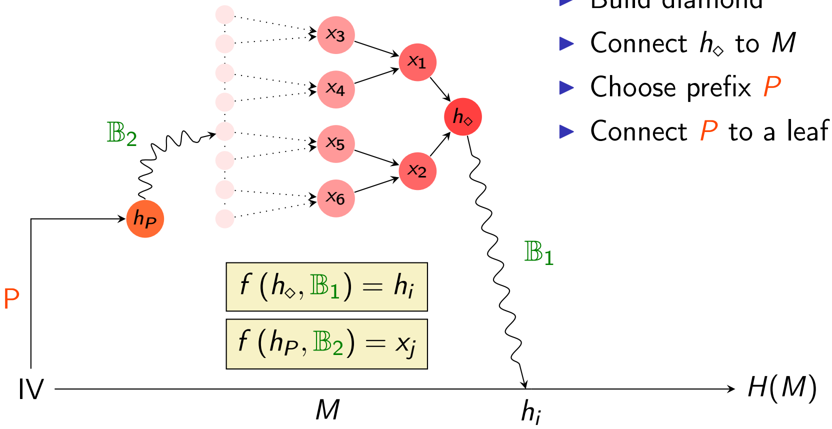
Replaying Kelsey and Schneier's attack, but with a diamond



- Build diamond
- Connect $h_\diamond$ to M
- Choose prefix P

Putting the "Diamond" at Work

Replaying Kelsey and Schneier's attack, but with a diamond



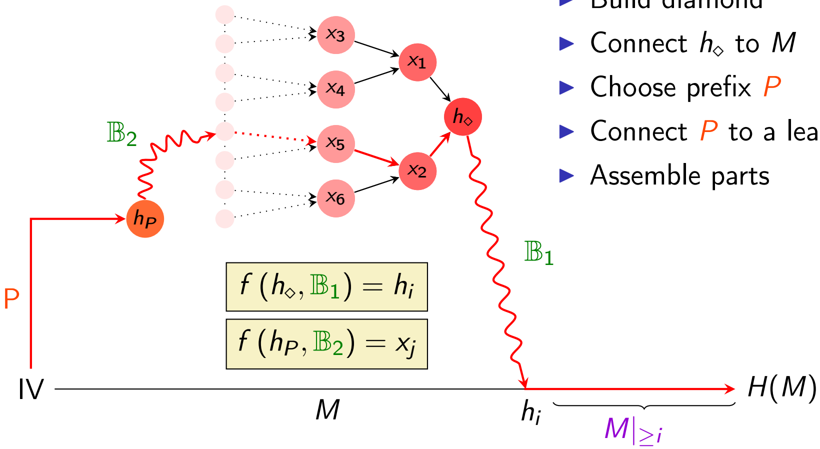
$$\boxed{f\left(h_\diamond, \mathbb{B}_1\right) = h_i}$$

$$\boxed{f\left(h_P, \mathbb{B}_2\right) = x_j}$$

- Build diamond
- Connect $h_\diamond$ to M
- Choose prefix P
- Connect P to a leaf x_j

Putting the “Diamond” at Work

Replaying Kelsey and Schneier’s attack, but with a diamond



- Build diamond
- Connect $h_\diamond$ to M
- Choose prefix P
- Connect P to a leaf x_j
- Assemble parts

Putting the "Diamond" at Work – Complexity

How much does this cost ? Assume $\left|M\right| = 2^k$.

1. Build diamond : $2^{n/2+\ell/2+2}$
2. Connect $h_\diamond$ to M : 2^{n-k}
3. Generate P : free
4. Connect h_P to Diamond : $2^{n-\ell}$
5. Assemble parts : free

Total : $2^{n/2+\ell/2+2} + 2^{n-k} + 2^{n-\ell}$

Take $\ell \simeq n/3$. Complexity becomes $\simeq 5 \cdot 2^{2n/3} + 2^{n-k}$

SHA-1 $(n = 160,\ k = 55,\ \ell = 53)$: complexity $= 2^{109.5}$

How To Cope With Rivest's Dithering ?

$$\mathbf{z} = abcacdcbcd\,cadcdbdabacabadbabcbdbcba\ldots$$

Question

How does this affect the attack ?

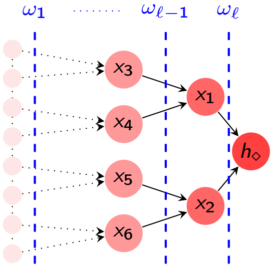
$\Longrightarrow$ We have to fix dithering symbols :

1. Inside the diamond
2. When connecting $h_\diamond$ to M

Key Ideas

- Fix a dithering symbol for each level of the diamond
  $\rightarrow$ ω_i at level i $\quad (1 \leq i \leq \ell)$
- guess the right symbol ($\omega_{\ell+1}$) for the connection

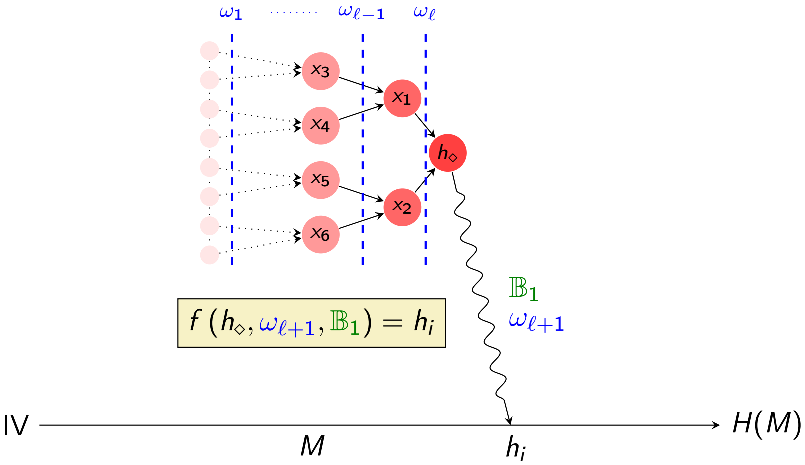
How To Cope With Rivest's Dithering (cont'd) ?



$$\mathrm{IV} \xrightarrow[M]{\hspace{10em}} H(M)$$

abcacdcbcdcadcdbdabacabadbabcbdbcbacbcdcacba . . .

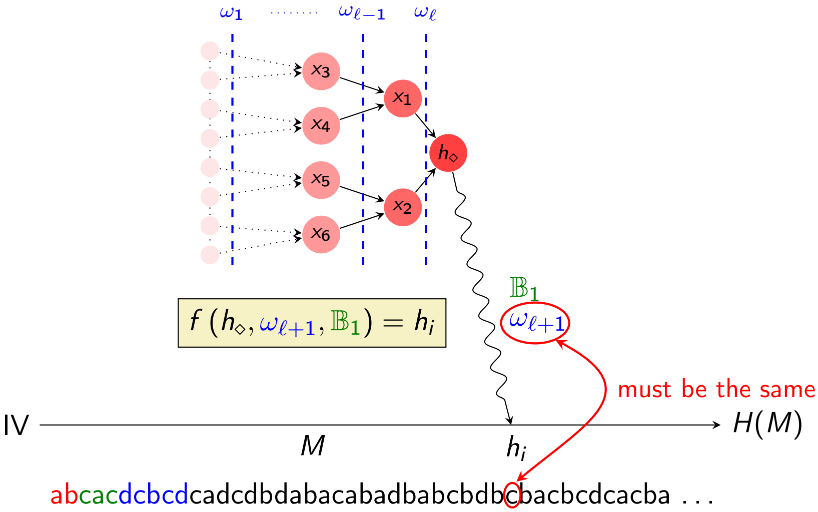
How To Cope With Rivest's Dithering (cont'd) ?



$$\boxed{f\left(h_\diamond, \omega_{\ell+1}, \mathbb{B}_1\right) = h_i}$$

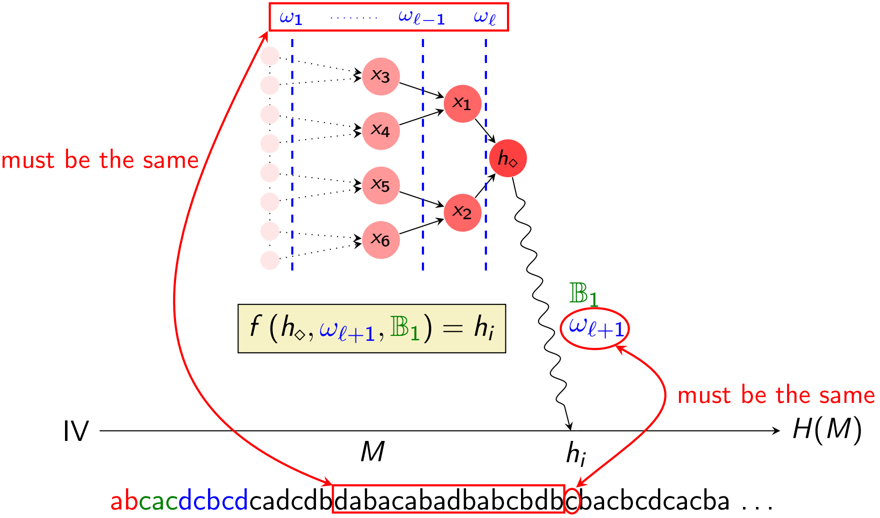
abcacdcbcdcadcdbdabacabadbabcbdbcbacbcdcacba . . .

How To Cope With Rivest's Dithering (cont'd) ?

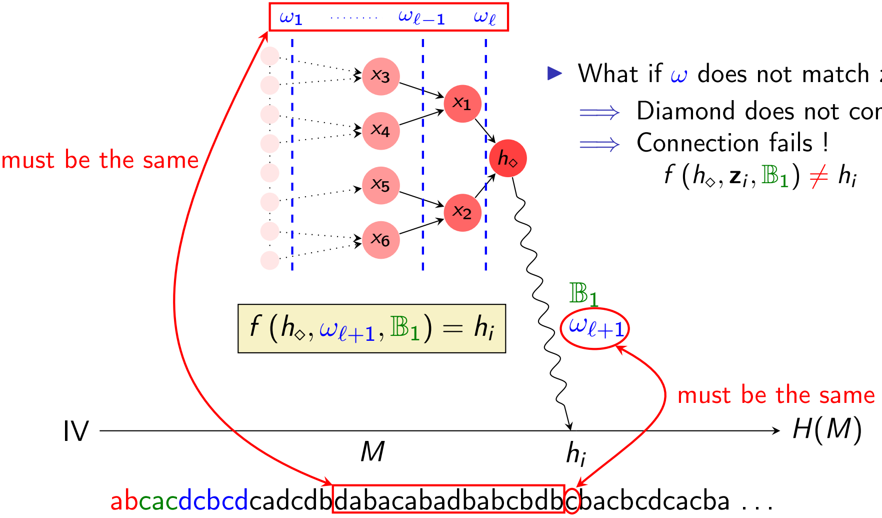


$$\boxed{f\left(h_\diamond, \omega_{\ell+1}, \mathbb{B}_1\right) = h_i}$$

How To Cope With Rivest's Dithering (cont'd) ?



How To Cope With Rivest's Dithering (cont'd) ?



$$\boxed{f(h_\diamond, \omega_{\ell+1}, \mathbb{B}_1) = h_i}$$

- What if ω does not match $\mathbf{z}$?

  $\Longrightarrow$ Diamond does not converge !

  $\Longrightarrow$ Connection fails !

  $$f(h_\diamond, \mathbf{z}_i, \mathbb{B}_1) \neq h_i$$

How To Cope With Rivest's Dithering ? (end)

With dithering, the diamond (and connection) only works at certain positions, where $\omega_{1\dots(\ell+1)}$ matches **z**.

Question

How to choose ω? Probability that ω matches **z** where $\mathbb{B}_1$ connects?

(Partial) Answer

Depends on **z**.

- Should choose a frequently-occuring factor of **z**
- Probability depends on how often it appears in **z**

Attack ?

Could there be frequently-occuring factors in **z** ?

Analysis of Rivest's dithering sequence
Or : How a Cryptanalyst Becomes a Sequence-Theorist for a While

Answer : YES

Theorem (Cobham,1972, "Uniform Tag Sequences")

The number of different factors of size s in $\mathbf{z}$ *is linear in s*

- There is a very low number of different factors in $\mathbf{z}$
  $\Longrightarrow$ so at least one of them occur frequently.
- Would have been exponential for a pseudo-random sequence...

Before, for SHA-1, we chose $\ell = 53$

- How many factors of size 54 in $\mathbf{z}$? 772 !
- Careful choice of ω:
  $\Longrightarrow$ Each connecting block $\mathbb{B}_1$ works with probability $\geq 2^{-9}$
  $\Longrightarrow$ Just repeat the attack 2^9 times !

Complexity

Same as before, except that many wrong connecting blocks $\mathbb{B}_1$ will be found before ω matches $\mathbf{z}$.

$$2^{n/2+\ell/2+2} + Fact_{\mathbf{z}}(\ell+1) \cdot 2^{n-k} + 2^{n-\ell}$$

For comparison with SHA-1, we take $n = 160$ and $k = 55$.

Hash function	ℓ	$Fact(\ell+1)$	SHA-1	Complexity
Plain-MD	55		$2^{109.5}$	$5 \cdot 2^{2n/3} + 2^{n-k}$
Keränen-Rivest	52	748	$2^{115.5}$	$(k+40.5) \cdot 2^{n-k+3}$
Concrete-Rivest	52	33176	2^{121}	2^{n-k+15}
Shoup's UOWHF	53	small	2^{112}	$(2k+3) \cdot 2^{n-k}$

- Keränen-Rivest is what was described before
- Concrete-Rivest is Rivest's "concrete proposal"
  (similar to Keränen-Rivest, but include a 13-bit counter)
- Shoup's UOWHF was presented at [EUROCRYPT'2000]

From One Long Message to Many Small Ones

Known generic second preimage attacks are long messages attacks

Possible to find a 2nd preimage of one out of many small messages

$h_{\mathcal{M}}$

- Connection step:
  - many small messages $\simeq$ one big message

  $\Rightarrow$ Target all of them at the same time

$$\mathrm{IV} \xrightarrow[M_1]{} H(M_1)$$

$$\mathrm{IV} \xrightarrow[M_2]{} H(M_2)$$

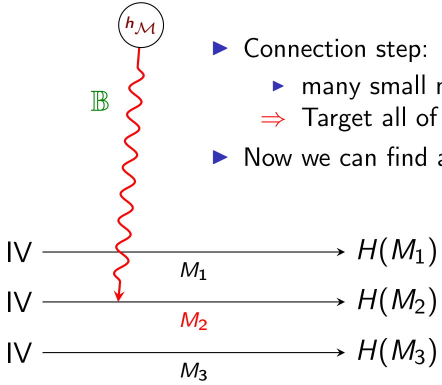
$$\mathrm{IV} \xrightarrow[M_3]{} H(M_3)$$

From One Long Message to Many Small Ones

Known generic second preimage attacks are long messages attacks

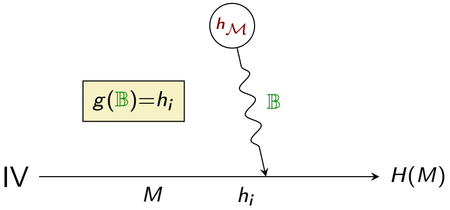
Possible to find a 2nd preimage of one out of many small messages

- Connection step:
  - many small messages $\simeq$ one big message
  - $\Rightarrow$ Target all of them at the same time
- Now we can find a second preimage of M_2 !



Faster Second Preimages With (quite a lot) More Precomputation

Hardest step : the connection. Let $g(\mathbb{B}) = f(h_{\mathcal{M}}, \mathbb{B})$.



- We need to find g^{-1} for one of the h_i
- Variation of Hellman's Time-Memory Tradeoff (2^n precomputation)
- Also works with shorter messages !

range of k	Memory	Time
$k \leq n/4$	$2^{2/3(n-k)}$	$2^{2/3(n-k)}$
$n/4 \leq k \leq n/2$	$2^{n/2}$	$2^{n/2}$

Conclusion

- New generic second preimage attack
  - About the first half of the preimage can be chosen
- Attack works in the presence of dithering
  - Rivest's proposal(s) are broken
  - First Attack on Shoup's UOWHF, ROX, ...
- Various extensions of both new and existing attacks
  - Apply attack to collection of small messages
  - Various possibilities for a Time-Memory Tradeoff
- Attack is not applicable to HAIFA...