# Security Analysis of Constructions Combining FIL Random Oracles

Yannick Seurin and Thomas Peyrin

France Télécom R&D and Université de Versailles

FSE ’07, March 26

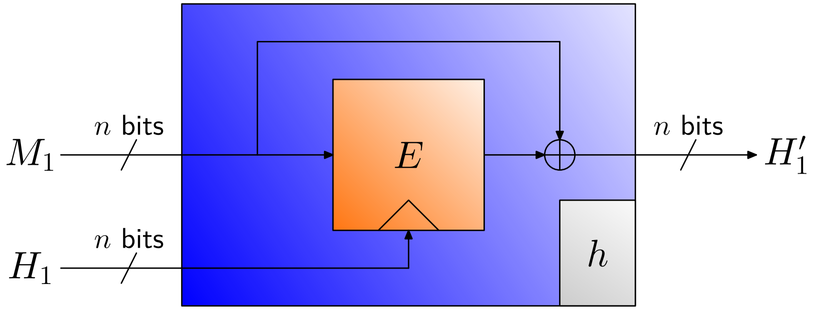
# Motivation: Block Cipher-Based Hash Functions

- Three well identified ways to design a compression function:
  - dedicated design (MD5, SHA-1, ...)
  - number theoretic design (VSH, MASH, ...)
  - block cipher-based design (Davies-Meyer, MDC-2, ...)

- “From scratch” compression functions come under attack
- Number theoretic designed hash functions suffer from poor performances
- ... so block cipher-based hash functions could be a promising way...

# Single vs. Multiple Block Length Hash Functions

- Single block length (SBL) hash functions are well understood since the work by Preneel *et al.* in 1993 and Black *et al.* in 2002, who provided security proofs in the ideal cipher model.
- Example: the Davies-Meyer construction (preimage resistance = $\Theta(2^n)$ queries, collision resistance = $\Theta(2^{n/2})$ queries)




- But single block length hash functions with 128-bits blocks block ciphers doesn't offer a sufficient security (brute force collision attacks need only $2^{64}$ work effort.)
- Therefore we need double (or multiple) block length hash functions in order to use AES for example.

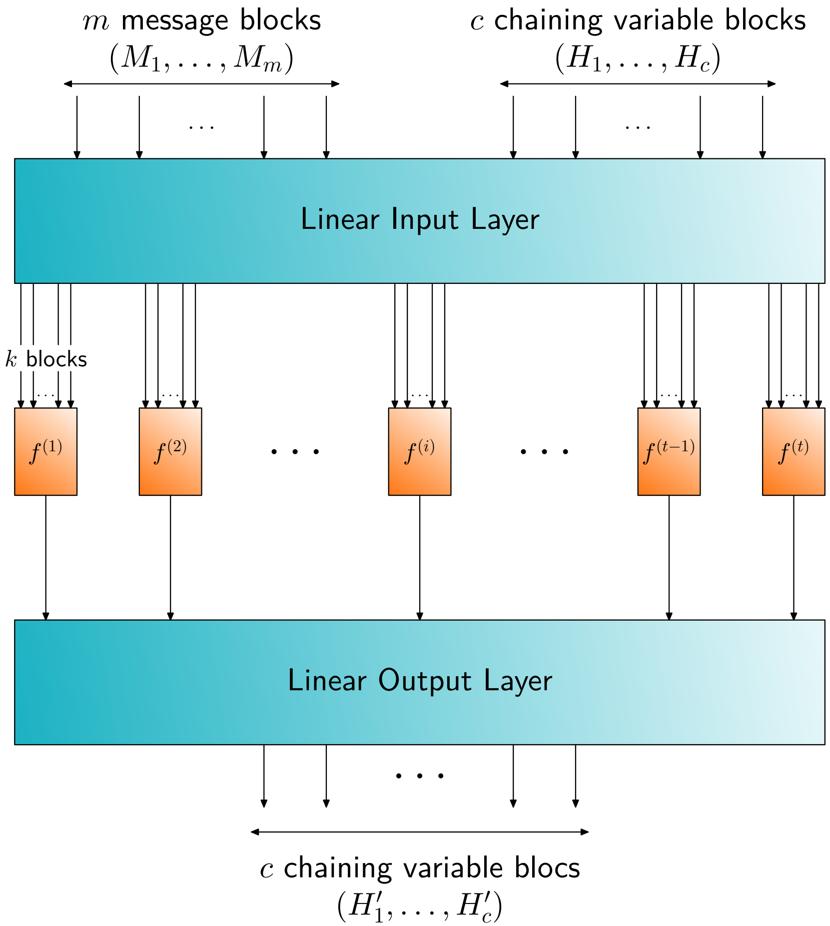
# Multiple Block Length Hash Functions

- No general theory for multiple block length hash functions as for SBL ones.
- A lot of candidate constructions have been proposed:
  - early proposals: ABREAST-DM, PARALLEL-DM, MDC-2, MDC-4
  - Knudsen-Preneel constructions (based on error correcting codes)
  - Hirose (FSE ’05, FSE ’06)
  - Nandi-Lee-Sakurai-Lee (FSE ’05)
- ... but very few remain unbroken.
- There is still no unbroken proposal of DBL hash function using a block cipher with key length equal to the block length (e.g. AES128).

# Our Contribution

- Recently Peyrin *et al.* [PGMR06] introduced a general framework for studying MBL hash functions and obtained necessary conditions for a MBL hash function to be secure by analysing generic attacks.
- They proved that a DBL compression function, using a block cipher with key length equal to the block length and hashing one or two blocks of message needs at least **five** independent block ciphers.
- They proposed new DBL hash functions constructions for which no attacks are known.
- However no security proofs were given.
- **We give a security analysis of their framework in the random oracle model, i.e. we give security bounds for preimage and collision resistance, and describe generic preimage and collision attacks which sometimes meet the security bound.**

# The Framework

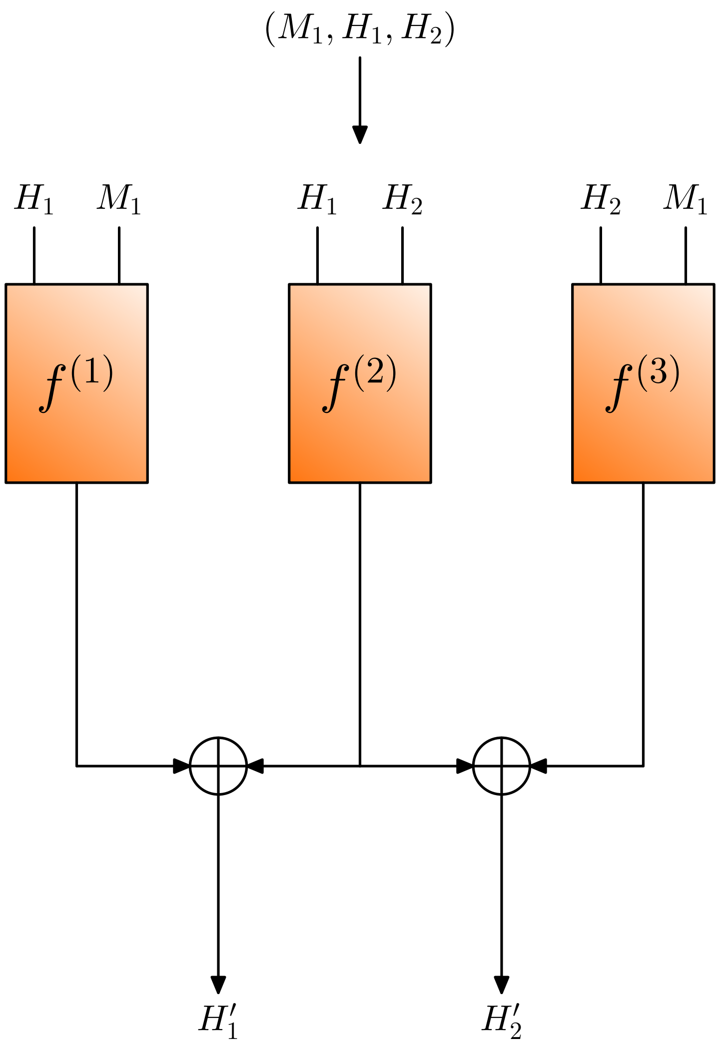



- We study generic constructions using:
  - $t$ compression functions $f_1, \ldots, f_t$
  - taking $k$ blocks of $n$ bits as input
  - outputting one block of $n$ bits
  - modelized as *independent random oracles*
- The resulting compression function:
  - takes $m$ message blocks of $n$ bits and $c$ chaining variable blocks of $n$ bits as input
  - outputs $c$ blocks of $n$ bits

# Computability Notions

- We will consider adversaries making at most $q$ queries to each inner compression function $f_1, \ldots, f_t$.
- We will need the following notions: let's fix sets of queries $Q_1, \ldots, Q_t$ to each inner compression function, and let's fix $r$ output blocks (or linear combination of output blocks) $(H'_{i_1}, \ldots, H'_{i_r})$. Then:
  - an input $(M_1, \ldots, M_m, H_1, \ldots, H_c)$ to the compression function $h$ is $(H'_{i_1}, \ldots, H'_{i_r})$-computable if the queries enable to compute the output blocks $(H'_{i_1}, \ldots, H'_{i_r})$
  - $\beta'_r(q)$ will be the maximum over the sets of queries and over the output blocks $(H'_{i_1}, \ldots, H'_{i_r})$ of the number of $(H'_{i_1}, \ldots, H'_{i_r})$-computable inputs.

# Computability Notions: Example




- Nandi *et al.* scheme N1 $(c = 2, m = 1, t = 3, k = 2)$.
- $\beta'_1(q) = q^2$
- Proof ($\geqslant$): fix $H_1$, choose $q$ values of $M_1$ and $H_2$, ask the $q$ queries $f_1(H_1, M_1)$ and $f_2(H_1, H_2)$. Then you can compute $H'_1$ for $q^2$ values $(M_1, H_1, H_2)$.
- $\beta'_2(q) \simeq q^{3/2}$
- Proof ($\geqslant$): choose $q^{1/2}$ values of $M_1$, $H_1$ and $H_2$, ask the $q$ queries $f_1(H_1, M_1)$, $f_2(H_1, H_2)$ and $f_3(H_2, M_1)$. Then you can compute $(H'_1, H'_2)$ for $(q^{1/2})^3$ values $(M_1, H_1, H_2)$.

# Generic Preimage Attacks

The following attack is a generalization of the Knudsen-Muller attack on the schemes of Nandi *et al.* and uses multipreimages on one output block (or linear combination of output blocks):

- choose the output block (or linear combination of output blocks) maximizing $\beta'_1(q)$ and compute the corresponding images for the output block
- for the inputs matching the preimage one is looking for, make the additional queries to compute the full image by $h$
- achieves advantage $\Omega\left(\frac{\beta'_1(q)}{2^{cn}}\right)$ as soon as $\beta'_1(q) = \Omega(n2^n)$.

# Generic Collision Attacks

We describe two possible collision attacks (which one is the better may depend of the construction):

- naïve one: compute $\beta'_c(q)$ hashes (advantage: $\Omega\left(\frac{\beta'_c(q)^2}{2^{cn}}\right)$ )
- multicollision on one output block:
  - choose the output block (or linear combination of output blocks) maximizing $\beta'_1(q)$ and compute the corresponding images for the output block
  - order the "collision classes" by decreasing order and look into them for a full collision
  - achieves advantage $\Omega\left(\frac{q\beta'_1(q)}{2^{cn}}\right)$ as soon as $\beta'_1(q) = \Omega(n2^n)$.

# Security Bounds

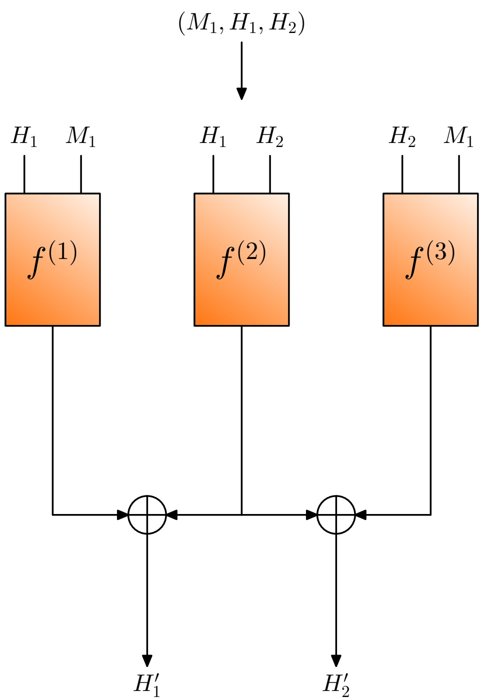
- We obtain the following bounds for the advantage of any adversary limited to $q$ queries:
  - $$\mathrm{Adv}_h^{\mathrm{pre}}(q) = O\left(\frac{\beta_1'(q)}{2^{cn}}\right)$$
  - $$\mathrm{Adv}_h^{\mathrm{coll}}(q) = O\left(\frac{\beta_1'(q)^2}{2^{cn}}\right)$$
- Idea of the proof: condition the probability of success of the adversary on the probability of success for **a single** output block.
- For the full proof, please see the paper.

# Summing Up the Results

| | Lower Bound | Upper Bound |
|---|---|---|
| Preimage Resistance | $\Omega\left(\frac{\beta_1'(q)}{2^{cn}}\right)$ | $O\left(\frac{\beta_1'(q)}{2^{cn}}\right)$ |
| Collision Resistance | $\Omega\left(\frac{\max(\beta_c'(q)^2, q\beta_1'(q))}{2^{cn}}\right)$ | $O\left(\frac{\beta_1'(q)^2}{2^{cn}}\right)$ |

- The analysis is tight in the case of preimage resistance: it is characterized by the parameter $\beta_1'(q)$.
- Things are more complex for collision resistance: the analysis is tight only in some particular cases.

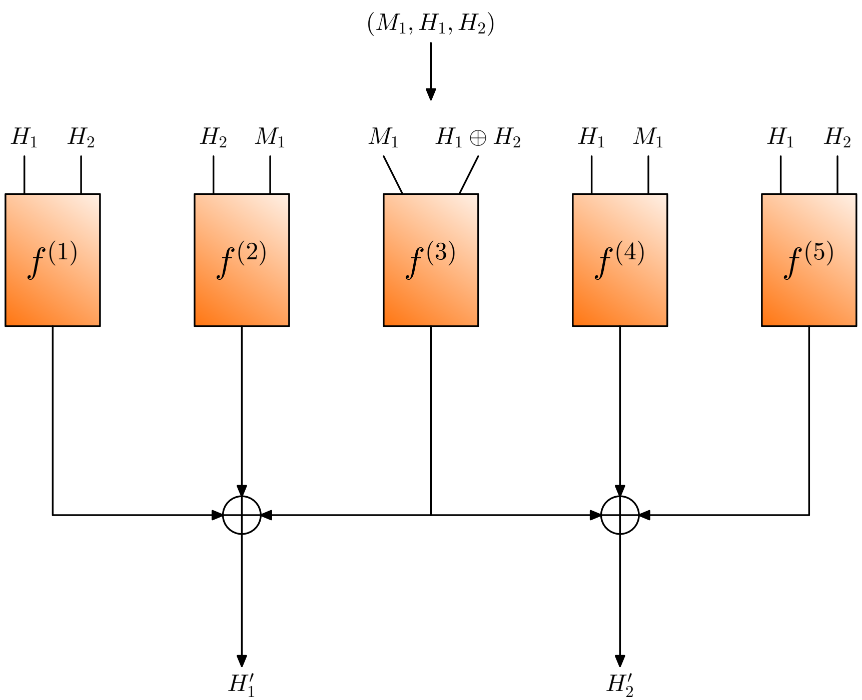
# Application to Previously Proposed Schemes




- Nandi *et al.* scheme N1
- For this scheme, $\beta'_1(q) = q^2$ and $\beta'_2(q) \simeq q^{3/2}$

| | Lower Bound | Upper Bound |
|---|---|---|
| Preimage Resistance | $\Omega\left(\frac{q^2}{2^{2n}}\right)$ | $O\left(\frac{q^2}{2^{2n}}\right)$ |
| Collision Resistance | $\Omega\left(\frac{q^3}{2^{2n}}\right)$ | $O\left(\frac{q^4}{2^{2n}}\right)$ |

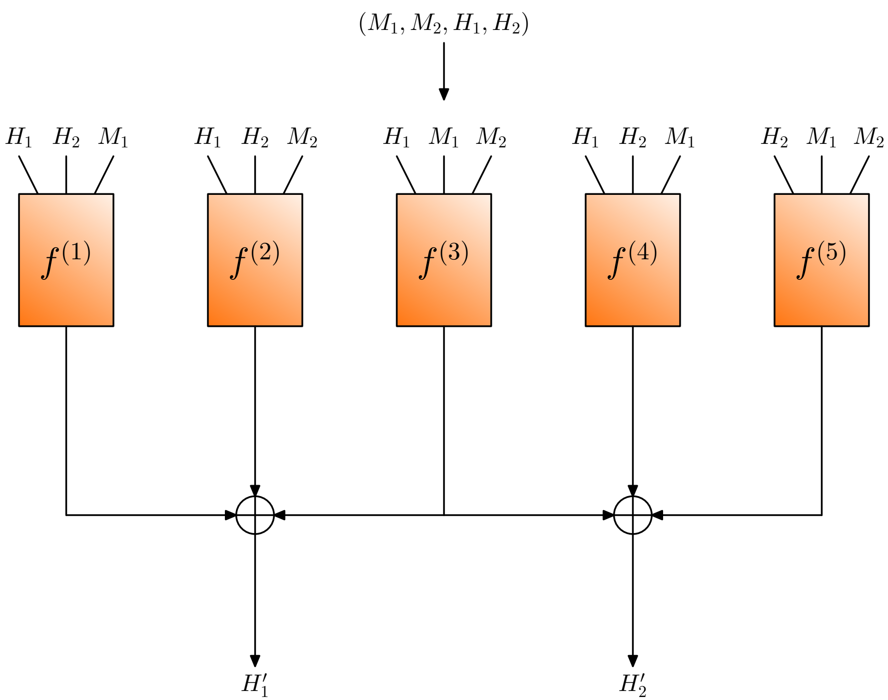
# Application to Previously Proposed Schemes




- Peyrin *et al.* scheme PGMR1
- For this scheme, $\beta'_1(q) \simeq q^{3/2}$ and $\beta'_2(q) \simeq q^{3/2}$

| | Lower Bound | Upper Bound |
|---|---|---|
| Preimage Resistance | $\Omega\left(\frac{q^{3/2}}{2^{2n}}\right)$ | $O\left(\frac{q^{3/2}}{2^{2n}}\right)$ |
| Collision Resistance | $\Omega\left(\frac{q^{3}}{2^{2n}}\right)$ | $O\left(\frac{q^{3}}{2^{2n}}\right)$ |

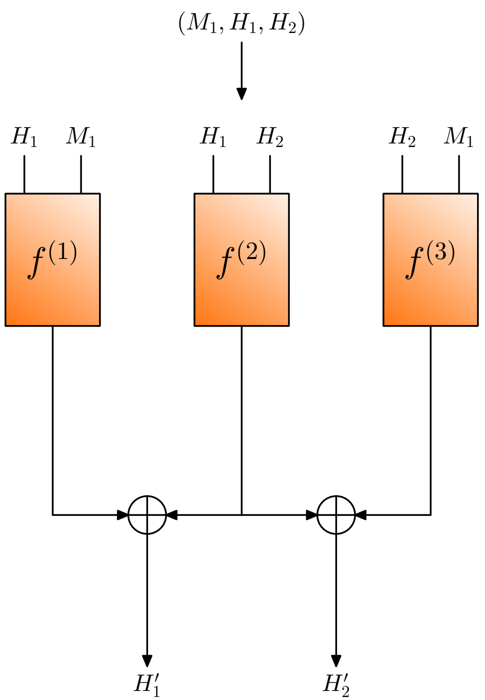
# Application to Previously Proposed Schemes




- Peyrin *et al.* scheme PGMR2
- For this scheme, $\beta'_1(q) \simeq q^{3/2}$ and $\beta'_2(q) \simeq q^{4/3}$

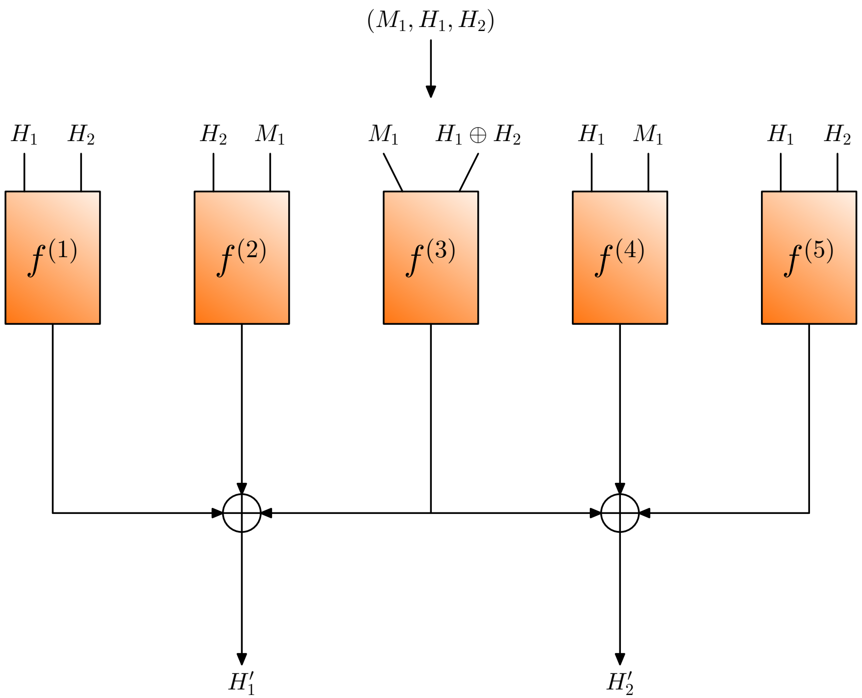
| | Lower Bound | Upper Bound |
|---|---|---|
| Preimage Resistance | $\Omega\left(\frac{q^{3/2}}{2^{2n}}\right)$ | $O\left(\frac{q^{3/2}}{2^{2n}}\right)$ |
| Collision Resistance | $\Omega\left(\frac{q^{8/3}}{2^{2n}}\right)$ | $O\left(\frac{q^{3}}{2^{2n}}\right)$ |

# Related Algorithmical Problems




- Distinction between security analysis in terms of **number of oracle queries** and **number of operations**.
- For this scheme, the preimage attack requires $O(2^n)$ queries and the collision attack requires $O(2^{2n/3})$ queries.
- But it is also possible to mount these attacks with resp. $O(2^n)$ and $O(2^{2n/3})$ **operations**.
- This is possible thanks to an efficient algorithm to solve the 2-sum problem...

# Related Algorithmical Problems




- Try to mount the multipreimage attack on this scheme (this requires $O(2^{4n/3})$ queries)...
- With $2^{4n/3}$ queries to $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$ you can obtain $2^{2n}$ images for $H_1'$. True...
- ...but how do you sort the ones which match the preimage you're looking for without effectively computing them (hence $2^{2n}$ operations...)?
- Strongly linked with the 3-sum problem...

# Conclusion and Future Work

- We studied the security of very general MBL hash function constructions in the FIL random oracle model.

- We gave security bounds for preimage and collision resistance and described generic preimage and collision attacks. Security analysis for preimage resistance is tight.

- Future work includes:
  - closing the security gap for collision resistance in terms of oracle queries
  - carrying out the analysis in the ideal block cipher model
  - understanding the security of (even basic) constructions in terms of computational complexity and the links with the $k$-sum problem.

# Thanks For Your Attention...

Questions?