Hash Functions from Defective Ideal Ciphers

Jonathan Katz, Stefan Lucks and
Aishwarya Thiruvengadam

CT-RSA 2015

Motivation

- Cryptographic constructions based on lower-level primitives are often analyzed by *modeling the primitive as an ideal object*
  - Sometimes, impossible to construct based on standard assumptions
  - Here: hash functions from block ciphers
- When instantiated, the primitive may have "defects" and be far from ideal

Motivating example

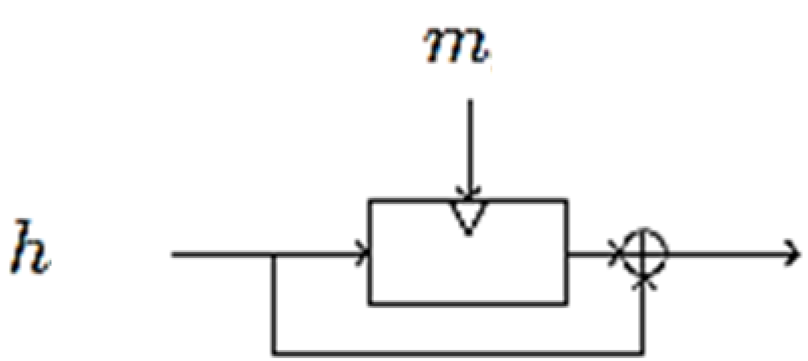
- *Related-key attacks on block ciphers*
  – Several such attacks on block ciphers are known
  – Does not contradict pseudorandomness
- Such attacks have been used to attack primitives based on (ideal) ciphers
  – Collision attack on the hash function used in Microsoft Xbox due to related-key attack on TEA
  – Attack on the RMAC message authentication code

This work

- We *define a "defective" ideal cipher model* incorporating linear related-key attacks
  - Goal: better understand real-world security of constructions analyzed in the (traditional) ideal-cipher model
- *We analyze the classical Preneel-Govaerts-Vandewalle (PGV) constructions* of hash functions from block ciphers in our model

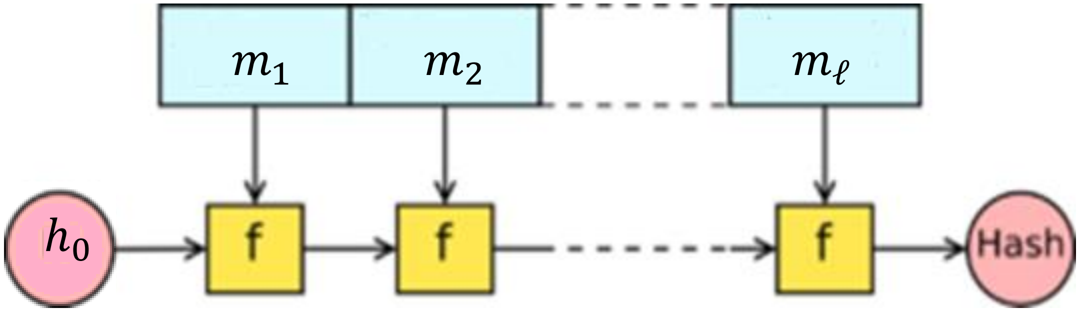
Background: Compression functions

- A (block-cipher-based) *compression function* $f: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ is a function that has oracle access to a block cipher $E: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$
  - For example, the Davies-Meyer compression function is defined as : $\mathrm{DM}(h, m) = E_m(h) \oplus h$



Iterated hash of compression function

- Let $f: \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$ be a (block-cipher-based) compression function and let $h_0 \in \{0,1\}^n$ be an arbitrary fixed constant.
- The Merkle-Damgard *iterated hash H* of the compression function f is defined as $H^f(m_1, \ldots, m_\ell) = h_\ell$ where $h_i = f^E(h_{\{i-1\}}, m_i)$



Hash functions and their security

- *Collision resistance* of block-cipher-based hash function H
  - Computationally unbounded adversary A given oracle access to E and E^{-1}
  - Adversary must make explicit and bounded number of queries to the oracle(s)
  - Aims to find a collision in H^E, i.e., messages $M \neq M'$ such that $H^E(M) = H^E(M')$
  - Security defined as the probability that A finds a collision where the probability is (also) taken over the choice of E.
- (Merkle-Damgard) Theorem : The hash function is collision-resistant if the underlying compression function is collision-resistant
  - Possible for hash function to be collision-resistant even if compression function is not

Results

- *None* of the PGV compression functions are collision-resistant in our “defective” ideal cipher model
- However, *four* of the PGV hash functions **are** collision-resistant in our model
  - In contrast to 20 collision-resistant PGV hash functions in the ideal-cipher model

Interpreting our results

- Our results do not imply anything about security of a specific instantiation
- But all else being equal, our results suggest using hash-function constructions robust to related-key weaknesses in the underlying cipher

Related work

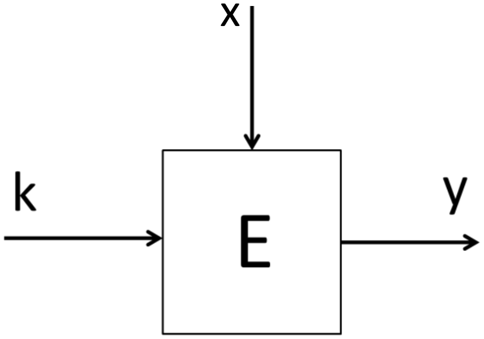
- Analysis of PGV functions in the ideal-cipher model [BRS02,BRSS10]
- Reducibility of block-cipher-based compression functions [BFFS13]
- “Weakened” random oracle models
  - Hash functions [Liskov06],Digital signature schemes [NIT08], Encryption schemes [KNTX10]
- Hash functions from weak compression functions [Lucks05]

Ideal cipher

- An *ideal cipher* is an oracle

  $$E: \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}^n$$

  where for each $k \in \{0,1\}^n$, the function $E_{k(\cdot)} = E(k,\cdot)$is chosen uniformly from the set of permutations on $\{0,1\}^n$.



Our model: Weakened ideal cipher

- Ideal *except* for the fact that the block cipher has *related-key weakness*
  - I.e., the block cipher returns related outputs on related keys/inputs.
- For a fixed key-shift $\Delta k \neq 0^n$ and fixed input-shift and output-shift $\Delta x, \Delta y \in \{0,1\}^n$:

$$\boxed{E_{\{k \oplus \Delta \mathrm{k}\}}(x \oplus \Delta x) \oplus \Delta y \coloneqq E_k(x)}$$

  - We exclude $\Delta k = 0^n$ because in that case E is not even pseudorandom.

Definition : Weakened ideal cipher

- Let $\Delta k \in \{0,1\}^n \backslash \{0^n\}$ and $\Delta x, \Delta y \in \{0,1\}^n$.
- Let $\mathrm{K} \subset \{0,1\}^n$ be such that $(\mathrm{K}, \mathrm{K} \oplus \Delta k)$ partitions {0,1}^{n}
- A *($\Delta k, \Delta x, \Delta y$)-ideal cipher* is an oracle $E: \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$
  - where for each $k \in K$, the function $E(k,\cdot)$ is uniform from the set of permutations on $\{0,1\}^n$
  - and for $k \notin \mathrm{K}$, we define $E_k(x) = E_{\{k \oplus \Delta \mathrm{k}\}}(x \oplus \Delta x) \oplus \Delta y$

Hash functions and their security

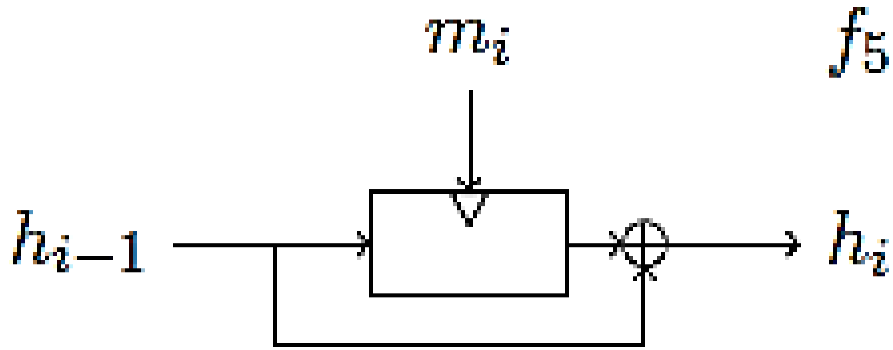
- Collision resistance of a hash function instantiated with a $(\Delta k, \Delta x, \Delta y)$-ideal cipher
  - Collision resistance definition as before but for block cipher E which is a $(\Delta k, \Delta x, \Delta y)$-ideal cipher
- *Collision resistance of a hash function instantiated with a weakened ideal cipher*
  - Collision resistant if: collision resistant with a $(\Delta k, \Delta x, \Delta y)$-ideal cipher for all values of $\Delta k \in \{0,1\}^n \setminus \{0^n\}$ and $\Delta x, \Delta y \in \{0,1\}^n$.

PGV constructions [Crypto '93]

- Defined 64 compression function constructions $f_i \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ for $i \in \{1, \ldots, 64\}$
- MD-iterated hash of the compression functions give hash functions H_i

Example: Davies-Meyer construction

- Definition : DM$(h, m) = E_m(h) \oplus h$



- Davies-Meyer compression function proven collision-resistant in the ideal-cipher model
- Notice that the key to the block cipher E is an input block

Collisions in Davies-Meyer

- Fix arbitrary Δk and $\Delta x, \Delta y = 0^n$
- Then, for $M = m$ and $M' = m \oplus \Delta k$, we have DM$(h, m) = E_m(h) \oplus h$ = $E_{\{m \oplus \Delta k\}}(h) \oplus h$ = DM(h, $m \oplus \Delta k) = DM(h, M')$
- Attack produces a collision in the Davies-Meyer hash function as well since
  - $H^E(m_1, \ldots, m_\ell) = H^E(m_1, \ldots, m_\ell \oplus \Delta k)$

Matyas-Meyer-Oseas (MMO) construction

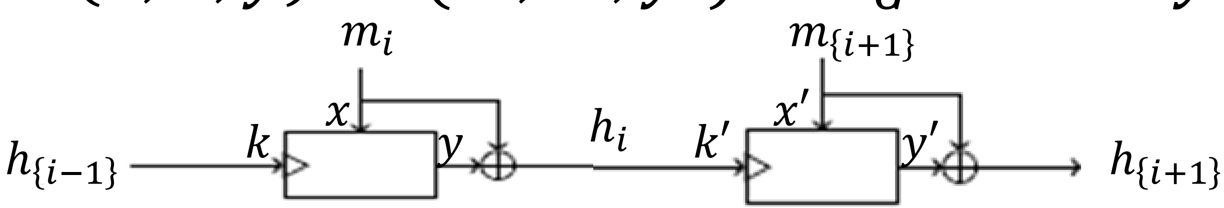
- Definition : $\text{MMO}(h, m) = E_h(m) \oplus m$
- Role of the chaining variable h and message m switched from Davies-Meyer
  - In particular, the key to the block cipher E does not depend on the input
- MMO compression function proven collision-resistant in the ideal-cipher model

Our result on MMO

- In our weakened ideal-cipher model, *the hash function is collision resistant* (but the compression function is not)
  - Recall that the compression function *is* collision-resistant in the ideal-cipher model

Collision resistance of MMO

- Define directed graph $G = (V_G, E_G)$
  - Vertex set $V_G = \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^n$
    - (x, k, y) denotes input, key and output of block cipher
    - If vertex (x, k, y) corresponds to round i of MMO, then $k = h_{\{i-1\}}, x = m_i$ and $h_i = E_{\{h_{\{i-1\}}\}}(m_i) \oplus m_i = E_k(x) \oplus x = y \oplus x$
    - If vertex (x', k', y') corresponds to round $i + 1$ of MMO, then $k' = h_i$
  - Arc $(x, k, y) \rightarrow (x', k', y')$ in E_G iff $k' = y \oplus x$



Collision resistance of MMO

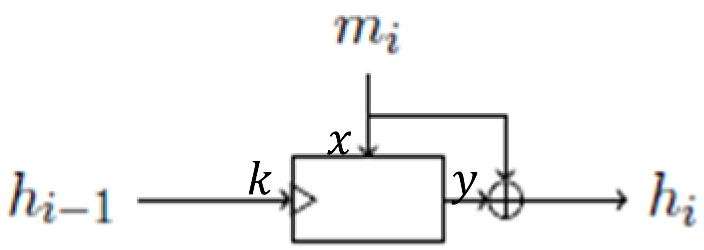
- Adversary A has access to E, E^{-1}oracles where E is a $(\Delta k, \Delta x, \Delta y)$-ideal cipher
- When A queries E on (k, x), oracle returns y in the form of the triple (x, k, y)
  - y chosen uniformly at random from the set of range points that have not been defined yet
  - The oracle also returns $(x \oplus \Delta x, k \oplus \Delta k, y \oplus \Delta y)$ (since A learns this by definition of $(\Delta k, \Delta x, \Delta y)$-ideal cipher)
- A's queries to E^{-1} are handled similarly

Collision resistance of MMO

- As A interacts with the oracle, color the vertices of the graph G as follows:
- When A asks an E-query, for each vertex returned,
  - If $k = h_0$, vertex (x, k, y) is colored red
  - Otherwise, vertex (x, k, y) is colored black

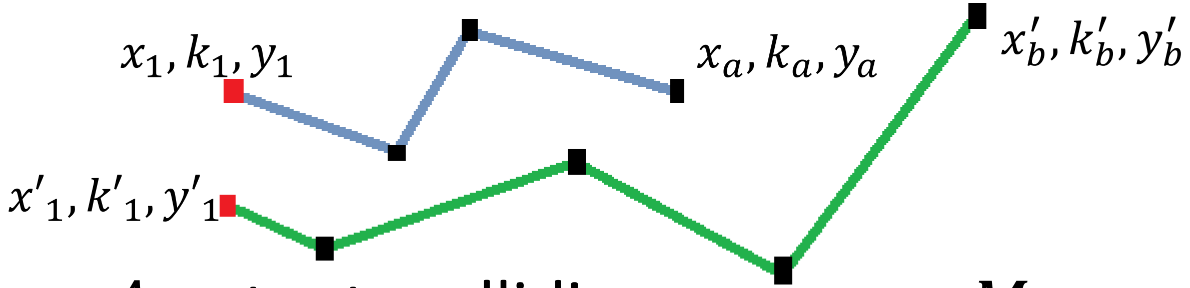
Collision resistance of MMO

- A *vertex of G is colored* if it gets colored red or black.
- A *path P in G is colored* if all of its vertices are colored.
- Vertices (x, k, y) and (x', k', y') *collide* if $y' \oplus x' = y \oplus x$.
- *Distinct paths P and P' are said to collide* if
  - All of their vertices are colored
  - Begin with red vertices
  - End with colliding vertices
- If A outputs two colliding messages, then there are necessarily two colliding paths.



Collision resistance of MMO:Proof

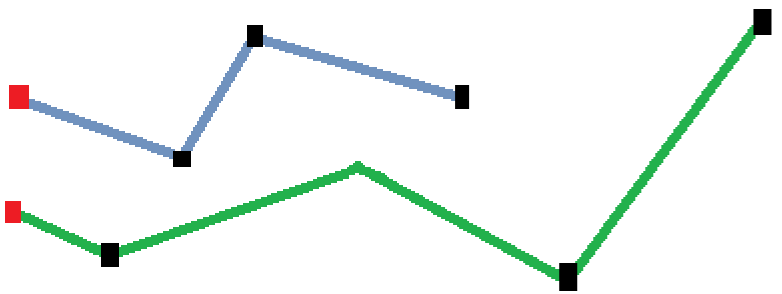
- Lemma : If A outputs two colliding messages, then there are necessarily two colliding paths.



- Suppose A outputs colliding messages $M = m_1 \ldots m_a$ and $M' = m'_1 \ldots m'_b$ such that $H^E(M) = H^E(M')$
- Let $P = (x_1, k_1, y_1) \rightarrow \cdots \rightarrow (x_a, k_a, y_a)$ where for each $i \in [a]$, $x_i = m_i, k_i = h_{\{i-1\}}, y_i = E_{\{k_i\}}(x_i)$ and $h_i = y_i \bigoplus x_i$. Define P' similarly. Then P and P' are colliding paths.

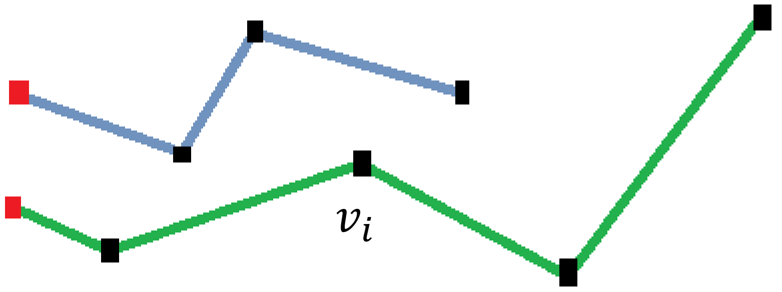
Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then
  - *A mid vertex got colored*



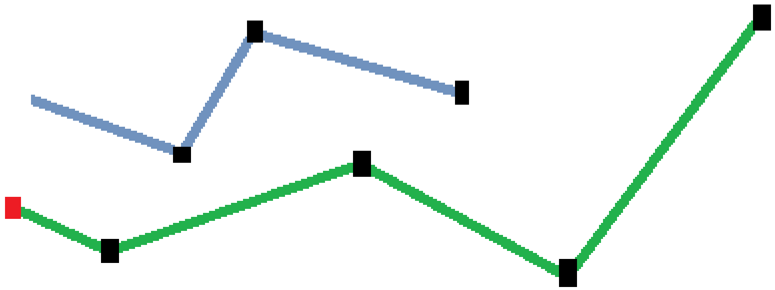
Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then
  - *A mid vertex got colored*



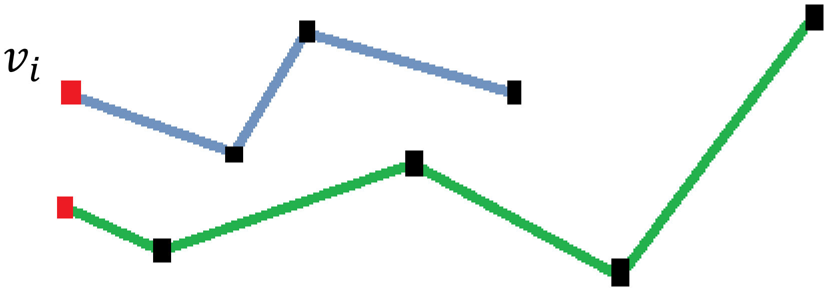
Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then
  - A mid vertex got colored or,
  - *A start vertex got colored*



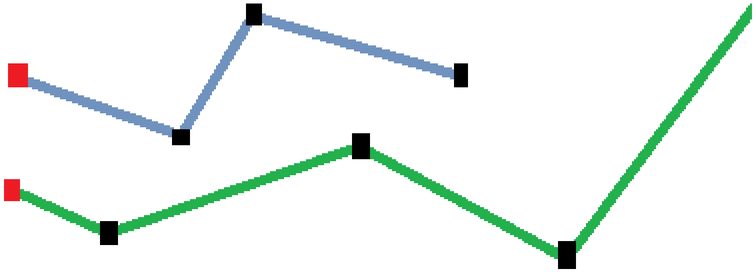
Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then
  - A mid vertex got colored or,
  - *A start vertex got colored*



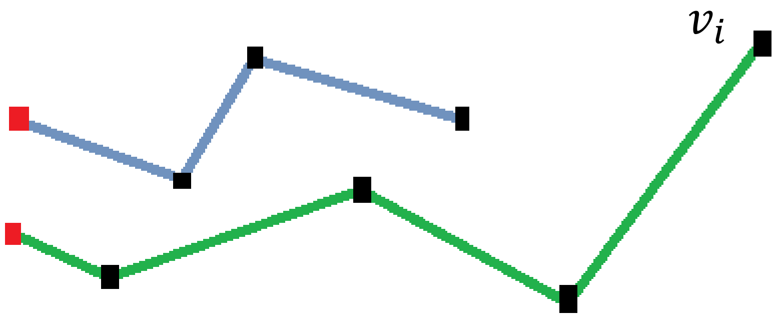
Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then
  - A mid vertex got colored or,
  - A start vertex got colored or,
  - *An end vertex got colored*



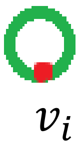
Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then
  - A mid vertex got colored or,
  - A start vertex got colored or,
  - *An end vertex got colored*



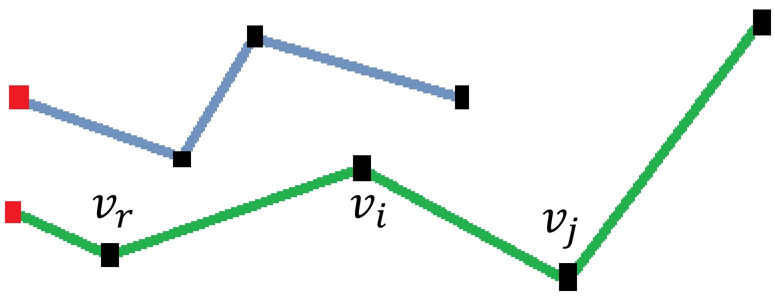
Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then
  - A mid vertex got colored or,
  - A start vertex got colored or,
  - An end vertex got colored or,
  - *A vertex colliding with itself got colored*



Collision resistance of MMO: Proof

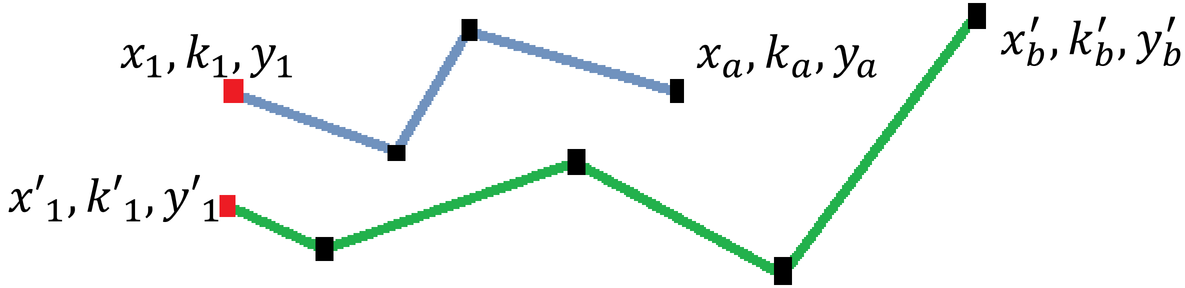
- If colliding paths are formed when the adversary asks query i (and not before) and a mid vertex v_i got colored
- Then, there exists vertices v_r and v_j which got colored in queries r and j such that there exists
  - Arc from v_r to v_i and
  - Arc from v_i to v_j i.e. $k_j = y_i \oplus x_i$



- Since either the x_i value or y_i value was chosen at random from a set of size at least $2^n - (i-1)$ and there are $2(i-1)$ possible options for v_j,
  - Prob(Arc from v_i to v_j) $\leq \{2(i-1)\}/\{2^n - (i-1)\}$
- There are 2 vertices returned for every query and it could so happen that both of these fall on a colliding path. In total, we get
  - Prob(a mid vertex gets colored) $\leq \{4(i-1)+2\}/\{2^n - (i-1)\}$

Collision resistance of MMO: Proof

- If colliding paths are formed when the adversary asks query i (and not before), then



- Analyzing all other cases similarly, we get
  - Prob(Colliding Paths) $\leq 14q(q+1)/2^n$, where q is the total number of queries made by A.

Conclusion

- Introduced a weakened ideal-cipher model
  - Meant to incorporate the possibility of related-key attacks (but no other structural weaknesses)
  - May be useful for analyzing other primitives as well
- Analyzed the PGV constructions in this model
- Proved that four PGV hash functions are collision-resistant up to the birthday bound in our model
- More results on inversion resistance and collision resistance of the rest of the hash functions

Thank you

RSA Conference 2015

San Francisco | April 20-24 | Moscone Center

CHANGE

Challenge today's security thinking

SESSION ID: CRYP-R02

Constructions of Hash Functions and Message Authentication Codes

Yusi Zhang

PhD in Computer Science
University of California, Davis
yzhangad@gmail.com

#RSAC

RSAConference2015

San Francisco | April 20-24 | Moscone Center

Use an Error-Correction Code for Fast, Beyond-birthday-bound Authentication

#RSAC

Motivation: Beyond-birthday-bound

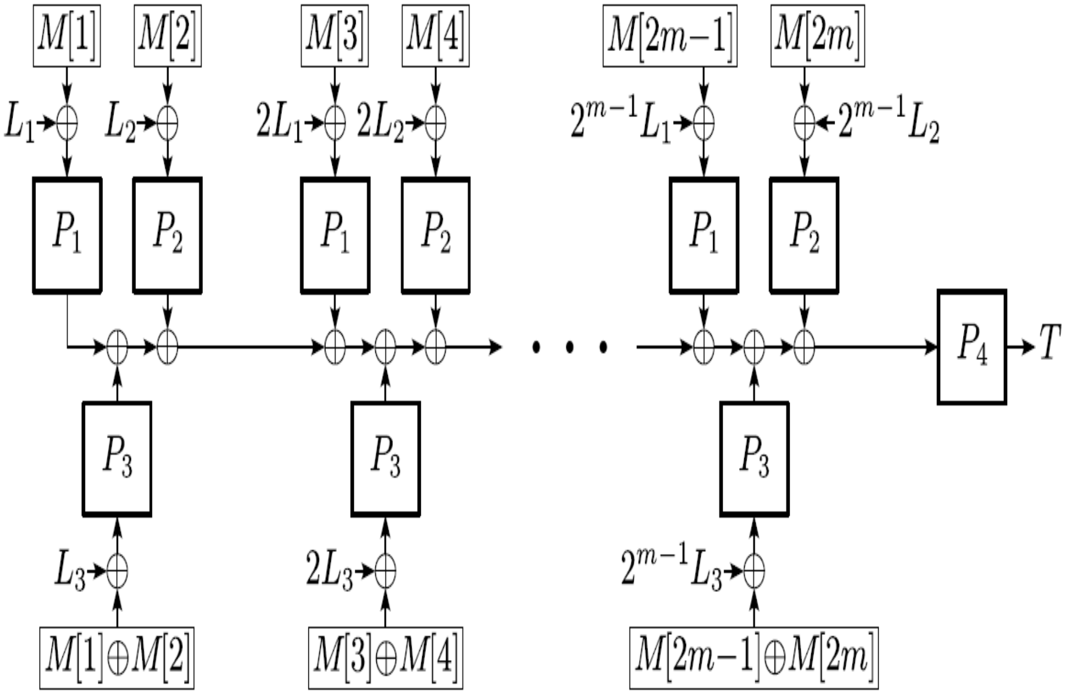
- Birthday Barrier: the $2^{n/2}$ - level.
- Best Known Bounds for Some MAC Modes:
  - CMAC: $O(q\sigma/2^n)$
  - PMAC: $O(q^2\rho/2^n)$
- Acceptable in Most Cases, but...

That depends on n!

Motivation Cont'd

- Problems:
  - Short 64-bit cipher is still widely deployed (financial institutions).
  - Hard to replace these ciphers (compatibility).
- Objective of this work:
  - Go beyond the Birthday Barrier.
  - Relatively Simple Modifications on an Existing Scheme (e.g. PMAC).
  - Avoid too much cost on efficiency and key setup.

Prior Work: PMAC with Parity (PMACwP) [Yasuda'12]

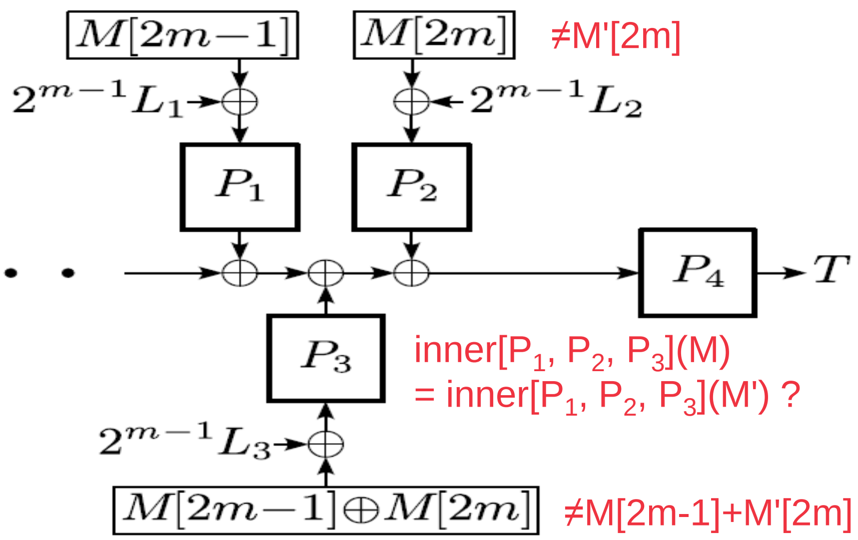


- Achieve a New Bound:

  $O(q^2/2^n + q\rho\sigma/2^{2n})$

- Shortcomings:
  - 4 independent keys needed.
  - 1.5 slowdown.

PMACwP: More Details about its Analysis



$$\Pr\Big[\texttt{inner}[P_1,P_2,P_3](M) = \texttt{inner}[P_1,P_2,P_3](M');\ P_1,P_2,P_3 \xleftarrow{\$} \mathrm{Perm}(n)\Big] \le \frac{1}{2^n} + \frac{m^2}{2^{2n}}$$

- Suffice to analyze the collision probability for the input to P_4.
- The $\mathbf{m^2/2^{2n}}$ term is the "source" of the beyond-birthday bound.
- Two key ingredients in the derivation to this term:
  - **Independence among the P_i's.**
  - **At least two different blocks.**
- **Will generalize, improve both.**

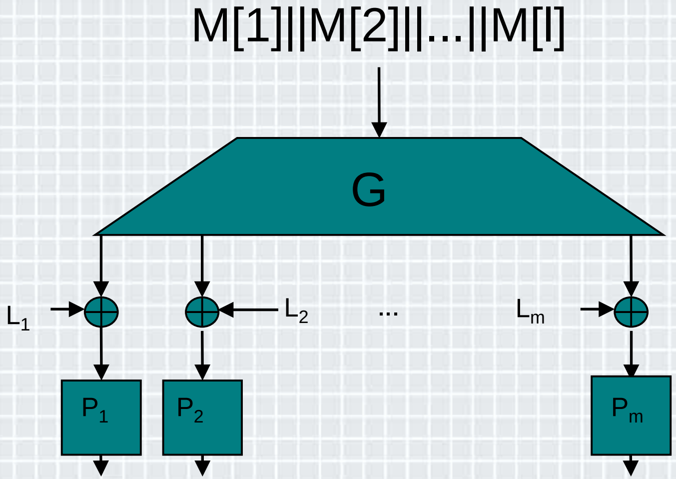
Generalization from 2 Differences to Multiple Ones

- M[1], M[2] -> M[1], M[2], M[1] + M[2] in matrix form:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

- What about a larger matrix?
- Desired Property: As many different output blocks as possible.
- Exactly the property of an MDS code.

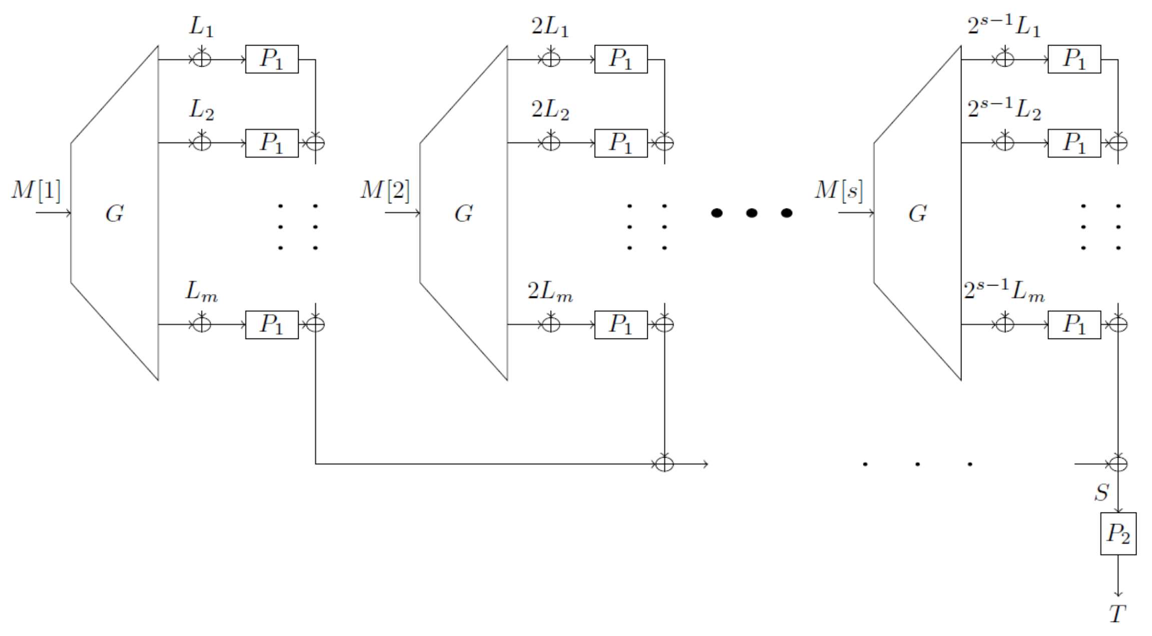
Generalization from 2 Differences to Multiple Ones



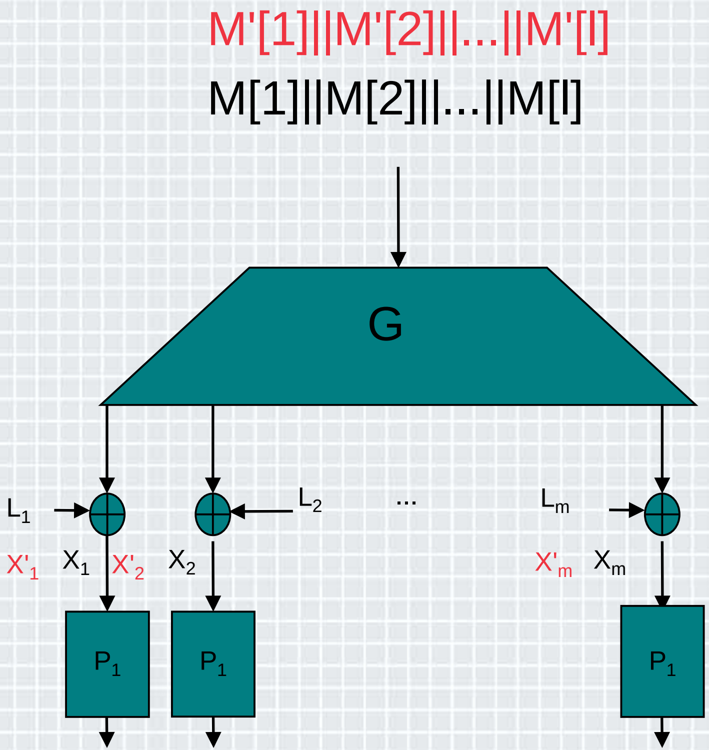
- Improve the bound to $O(q^2/2^n+q\sigma\rho^{d-1}/2^{dn})$
- But even more keys are needed...

Reduce the Number of Keys

- In the analysis, only interested in the collision of the final input.
- Possible to replace the many independent ciphers with a single one.
- Of course, a new proof becomes necessary...



Key Step in Our New Analysis



- $L_1, L_2, \ldots, L_m$ are randomly chosen.
- M, M' are fixed, with some difference in the first unit.
- Suppose every input to P_1 has been computed, except the red ones.
- Bad event in interest:

All the red X's collide with some previous inputs.

Key Step in Our Analysis, cont'd

- The MDS property excludes the trivial collision: $X_1 = X'_1$.
- If we fix the index of collided inputs, the event can be described by a matrix equation.

$$A \cdot L = B$$

An m-row matrix, each row encoding a collision and containing at most two non-zero entries.

The column vector: $[L_1, L_2, ..., L_m]^T$

The difference vector, depending only on M and M', hence a fixed vector.

The probability that this equation holds depends on the rank of A.

Key Step in Our Analysis, cont'd

- In general, the rank of A is unknown.
- However, among the m subkeys, at least half of them collide with subkeys of larger or equal indexes.
- Hence, if we focus only on such subkeys, we have a submatrix of A that is in row echelon form, therefore full-rank.
- The halving of A degrades the bound from $O(q^2/2^n+q\sigma\rho^{d-1}/2^{dn})$ to $O(q^2/2^n+q\sigma\rho^{(d-1)/2}/2^{(d+1)/2})$.
- **But, we've reduced the key number from m+1 to 2 only!**

Summary

- We've generalized Yasuda's PMACwP by introducing an MDS matrix into its preprocessing stage.
- Based on the basic generalization, we further reduced the number of keys to 2, at the cost of a degradation of provable security.
- Theoretically, our scheme can achieve a rate arbitrarily close to 1, a security level arbitrarily close to 2^n, by choosing large enough MDS matrices.
- Surprisingly, the above can be done by 2 independent keys only.

Candidate Topics for Future Work

- Reduce the number of keys even further: 2 to 1?
- Go beyond "birthday-barrier" for query numbers, q, as well.
- Analysis of Online Security.