# Methods for Collisions in some Algebraic Hash Functions

Simran Tinani

University of Zurich UZH

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

June 19, 2023

## Hash Functions

$\mathcal{A}$: alphabet; $\mathcal{A}^*$: all finite-length words in $\mathcal{A}$; $\mathcal{A}^n$: words up to length $n$ in $\mathcal{A}$.

Example: $\mathcal{A} = \{0, 1\}$; $\mathcal{A}^2 = \{0, 1, 00, 11, 01, 10\}$;
$\mathcal{A}^* = \{0, 1, 00, 11, 01, 10, 000, 111, 001, 010, 100, 011, 101, 110, ...\}$.

### Definition

*A length $n$ hash function, or compression function, is a map $\mathcal{A}^* \to \mathcal{A}^n$. A hash function $h : \mathcal{A}^* \to \mathcal{A}^n$ is called a cryptographic hash function if it satisfies the following properties:*

- ***Collision-resistance**: it is computationally infeasible to find a pair $x, x'$ of distinct messages such that $h(x) = h(x')$.*
- ***Second pre-image resistance**: given a message $x$, it is computationally infeasible to find another message $x' \neq x$ such that $h(x) = h(x')$.*
- ***One-wayness**: given a hash value $y \in \mathcal{A}^n$ it is computationally infeasible to find a pre-image $x \in \mathcal{A}$ such that $h(x) = y$.*

## Cayley Hash Functions

$G$: finite group with generator set $S = \{s_1, \ldots, s_k\}$; $|\mathcal{A}| = |S|$, so there is an associative binary operation

$$\star : G \times G \to G$$

and every element $g \in G$ has an expression $g = s_1^{e_1} \ldots s_k^{e_k}$.

**Definition (Cayley hash function)**

*Given an injective map $\pi : \mathcal{A} \to S$, define the hash value of the message $x_1 x_2 \ldots x_k$ to be the group element $\pi(x_1)\pi(x_2)\ldots\pi(x_k)$.*

Security $\equiv$ some concise mathematical problem; inherently parallelizable.

**Definition (Factorization problem)**

*Let $L > 0$ be a fixed constant. Given $g \in G$, return $m_1, \ldots, m_L$ and $\ell \leq L$, with $m_i \in \{1, \ldots, k\}$ such that $\prod_{i=1}^{\ell} s_{m_i} = g$.*

# Motivation

- Several widely used hash functions, including the NIST-standardized SHA (Secure Hash Algorithms) functions, are built from block ciphers.
- The security of block cipher-based hash functions is heuristic: it does not reduce to a well-known difficult mathematical problem.
- The security of Cayley hash functions is equivalent to some concise mathematical problem, i.e. "provable security".
- Cayley hashes are inherently parallelizable, i.e. allow simultaneous computation of the hash value of different parts of the message, and recombining these at the end.
- However, Cayley hash functions are also inherently mallaeble: given a hash $h(m)$ of an unknown message $m$, $h(x_1||m||x_2) = h(x_1)h(m)h(x_2)$ for any texts $x_1, x_2$.
- Further, they lack preimage resistance for small messages. However, there exist fixes for these disadvantages.

# Famous Cayley Hash Functions

$SL_2(\mathbb{F}_{p^k})$: Special $2 \times 2$ matrix group over finite field $\mathbb{F}_{p^k}$

**Definition (Zémor Hash Function, (Zémor, 1991))**

*For generators* $A_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ *of* $SL_2(\mathbb{F}_p)$, *a message* $m = m_1 m_2 \dots m_k \in \{0,1\}^*$ *define* $H(m_1 \dots m_k) = A_{m_1} \dots A_{m_k}$.

**Definition (Tillich-Zémor Hash function)**

*Let* $n > 0$ *and* $q(x)$ *be an irreducible polynomial over* $\mathbb{F}_2$. *Write* $K = \mathbb{F}_2[x]/q(x)$. *Consider* $A_0 = \begin{pmatrix} x & 1 \\ 1 & 0 \end{pmatrix}$ *and* $A_1 = \begin{pmatrix} x & x+1 \\ 1 & 1 \end{pmatrix}$, *which are generators of* $SL_2(K)$. *For a message* $m = m_1 m_2 \dots m_k \in \{0,1\}^*$ *define* $H(m_1 \dots m_k) = A_{m_1} \dots A_{m_k} \pmod{q(x)}$.

# Generalizations of Algebraic Hash Functions

**Definition (Generalized Zémor Hash Functions)**

*Consider the generators* $A_0 = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$ *and* $A_1 = \begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix}$ *in the group* $SL_2(\mathbb{F}_{p^k})$. *For a message* $m = m_1 m_2 \dots m_k \in \{0,1\}^*$ *define the hash value* $H(m_1 \dots m_k) = A_{m_1} \dots A_{m_k}$.

$A_0$, $A_1$ have order $p$, so one trivially has collisions of length $p$ with the empty word. Want to find collisions with length at most, say $\mathcal{O}(\sqrt{p})$.

**Definition (Generalized Tillich-Zémor hash functions)**

*Consider the generators* $A_0 = \begin{pmatrix} \alpha & 1 \\ 1 & 0 \end{pmatrix}$ *and* $A_1 = \begin{pmatrix} \beta & 1 \\ 1 & 0 \end{pmatrix}$ *where* $\alpha, \beta \in \mathbb{F}_{p^k}$, *in the group* $SL_2(\mathbb{F}_{p^k})$. *For a message* $m = m_1 m_2 \dots m_k \in \{0,1\}^*$ *define the hash value* $H(m_1 \dots m_k) = A_{m_1} \dots A_{m_k}$.

# Collisions from Triangular and Diagonal Matrices

- (Petit et al., 2009): if one can produce "sufficiently many" messages whose images in the matrix groups are upper/lower triangular, then one can find collisions of the generalized Zémor and Tillich-Zémor hash functions. Find $m$ such that

$$h(m) = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}; h(m) = \begin{pmatrix} a & b \\ c & 0 \end{pmatrix}; h(m) = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$$

**Problem 1 (Triangularising Hashes)**

*Given a matrix $C \in SL_2(\mathbb{F}_{p^k})$ formed as product of $A_0$ and $A_1$, find the conditions under which there exist integers $m$ and $n$ (of size significantly smaller than $p^k$) such that $CA_0^m A_1^n$ is upper/lower triangular, or even diagonal. Compute $m$ and $n$ if they exist.*

# Generalized Zémor hash functions

# Extending Messages for Triangular Zémor Hashes

**Lemma 1**

*Let $k \geq 1$ and $\alpha \cdot \beta \in \mathbb{F}_p$. Let $z$ be any message and $C := H(z)$ be its corresponding hash value. Assume that $a := C[0,0] \neq 0$. Then, there exist integers $m, n \in \{0, 1, \ldots, p-1\}$ such that $C \cdot A_0^m \cdot A_1^n$ is upper triangular.*

**Proposition 1**

*If $\alpha \cdot \beta \notin \mathbb{F}_p$, then $C \cdot A_0^m \cdot A_1^n$ is upper triangular for $m, n \in \mathbb{F}_p$ if and only if for*

$$\gamma = \left( \frac{d((d\beta)^{p-1} - c^{p-1})}{\alpha c^p (1 - (\alpha\beta)^{p-1})} \right), \tag{1}$$

*we have $\gamma^p = \gamma$, and $m = \gamma$; $n = \dfrac{-c}{\beta(mc\alpha + d)}$. If $k = 2$ then $\gamma^p = \gamma$ always holds.*

# Condition for Triangularisability

Can we generalize this method to make $C \cdot A_0^{m_1} A_1^{n_1} \dots A_0^{m_r} A_1^{n_r}$ upper/lower triangular and thereby extend the result to all $SL_2(\mathbb{F}_{p^k})$? For an extension where multiplication by a product $A_0^m A_1^n$ is allowed twice:

**Lemma 2**

*For $C := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, there exits integers $m_1, m_2, n_1, n_2$ such that $CA^{m_1}B^{n_1}A^{m_2}B^{n_2}$ is upper triangular if and only if the equation*

$$q_3 x^2 y + q_2 xy + q_1 y + q_0 = 0 \tag{2}$$

*has a solution $(x, y) \in \mathbb{F}_p \times \mathbb{F}_p$, where $q_0, q_1, q_2, q_3$ are given by*

$$\begin{aligned} q_3 &= c^{p^2} \alpha\beta((\alpha\beta)^{p^2-1} - 1), \\ q_2 &= c^{p^2} \gamma\alpha\beta(\gamma^{p-1} - (\alpha\beta)^{p^2-1}) + d\beta((d\beta)^{p^2-1} - 1), \\ q_1 &= d\beta\gamma(c^{p^2}\gamma^{p-1} - (d\beta)^{p^2-1}), \\ q_0 &= c^{p^2}\gamma(\gamma^{p-1} - 1). \end{aligned} \tag{3}$$

## Example: Condition for Triangularisability

### Example 1

For simplicity, consider the field $\mathbb{F}_{2^5}$ with generator $z_5$ and $\alpha = z_5^3 + 1$, $\beta = z_5^3 + z_5^2 + 1$. Consider the hash matrix

$$C = \begin{pmatrix} z_5^4 + z_5^3 + z_5^2 + z_5 & z_5^4 + z_5^3 + z_5^2 + z_5 \\ z_5^3 & z_5^4 + z_5^3 + z_5^2 \end{pmatrix}.$$

Here, we have $\gamma = z_5^4 + z_5 + 1$ and the polynomial in Equation (2) is $(z_5^2 + z_5)x^2y + (z_5^3 + z_5^2 + 1)xy + (z_5^3)y + (z_5^4 + z_5^2 + z_5)$. The $\langle (z_5^2 + z_5)x^2y + (z_5^3 + z_5^2 + 1)xy + z_5^3 y + (z_5^4 + z_5^2 + z_5), x^p - x, y^p - y \rangle$ is trivial, so its Gröbner basis is $\{1\}$. So, no solution exists.

## Example: Collisions

For $p = 7919$, $\alpha = 5698$, $\beta = 6497$, consider the message text

$$\begin{aligned} z = & 0^{44}1^{41}0^{17}1^{49}0^{47}1^{17}0^{50}1^{31}0^{15}1^{10}0^{39}1^{12}0^{2}1^{0}0^{24}1^{41}0^{28}1^{23}0^{9}1^{0}0^{47}1^{23}0^{1}1^{30}0^{18} \\ & 1^{32}0^{24}1^{14}0^{0}1^{49}0^{19}1^{28}0^{24}1^{26}0^{26}1^{26}0^{11}1^{1}0^{17}1^{20}0^{38}1^{22}0^{12}1^{38}0^{8}1^{33}0^{39}1^{42}0^{47}1^{29} \\ & 0^{10}1^{41}0^{14}1^{45}0^{13}1^{40}0^{42}1^{13}0^{2}1^{6}0^{40}1^{31}0^{2}1^{27}0^{1}1^{7}0^{36}1^{19}0^{3}1^{25}0^{10}1^{27}0^{21}1^{2}0^{12}1^{23} \\ & 0^{36}1^{8}0^{25}1^{39}0^{36}1^{0}0^{19}1^{39}0^{37}1^{32}0^{14}1^{4}0^{3}1^{12}0^{16}1^{23}0^{49}1^{25}0^{23}1^{19}0^{46}1^{23}0^{36}1^{31} \end{aligned}$$

We have, $H(z) = \begin{pmatrix} 4812 & 5537 \\ 4987 & 1690 \end{pmatrix} \in SL_2(\mathbb{F}_p)$.

Then for $z_1 = 1^{30} \cdot z \cdot 0^{6226}1^{744}$ and $z_2 = 1^{33} \cdot z \cdot 0^{6226}1^{180}$ we have the collision $H(z_1) = H(z_2) = \begin{pmatrix} 4812 & 0 \\ 0 & 1542 \end{pmatrix}$.

# Generalized Tillich-Zémor hash functions

## Generalized Tillich-Zémor Hash Functions

Consider the generalized Tillich-Zémor hash function $\phi$ with the generators $A_0 = \begin{pmatrix} \alpha & 1 \\ 1 & 0 \end{pmatrix}$ and $A_1 = \begin{pmatrix} \beta & 1 \\ 1 & 0 \end{pmatrix}$ where $\alpha, \beta \in \mathbb{F}_{p^k}$.

Consider the matrix $Y = \begin{pmatrix} x & 1 \\ 1 & 0 \end{pmatrix}$ and first compute its powers.

$$Y^n = \begin{pmatrix} f_n(x) & f_{n-1}(x) \\ f_{n-1}(x) & f_{n-2}(x) \end{pmatrix}, \; n \geq 2 \tag{4}$$

where $f_0(x) = 0$, $f_1(x) = 1$, and

$$f_n(x) = x f_{n-1}(x) + f_{n-2}(x) \tag{5}$$

It is clear that the recurrence relation (5) fully describes the powers of the matrix $Y$.

## Computing $f_n(x)$ for characteristic $p \neq 2$

We may solve (5) by finding roots of the auxiliary polynomial $t^2 - xt - 1 = 0$. It can be shown that for any $n \geq 1$, we have

$$f_n(x) = \frac{1}{2^{n+1}} \left[ \sum_{0 \leq i \leq n,\ n-i \text{ is even}} \sum_{j=0}^{(n-i)/2} \binom{n+1}{i} \binom{(n-i)/2}{j} 2^{n-2j} x^{i+2j} \right] \in \mathbb{F}_p[x]$$

Powers of $A_0$ and $A_1$ may therefore be computed in constant time.

## Condition for Collisions

- $\mathbb{F}_{p^k}$ is viewed through the isomorphism $\mathbb{F}_{p^k} \cong \mathbb{F}_p[x]/\langle q(x) \rangle$ where $q(x)$ is an irreducible polynomial of degree $k$ over $\mathbb{F}_p$.
- Thus, $\gamma \in \mathbb{F}_{p^k}$ is a polynomial of degree smaller than $k$, say $\gamma = g_\gamma(x)$.
- $f_n(\gamma)$ can be calculated as a polynomial modulo $q(x)$ by simply composing $f_n$ and $g$, i.e. $f_n(\gamma) = f_n(g_\gamma(x)) \pmod{q(x)}$.

**Lemma 3**

*Suppose that the adversary can compute integers $m$ and $n$ such that $f_{n-1}(g_\alpha(x)) = f_{m-1}(g_\beta(x)) \pmod{q(x)}$ and $f_{n-2}(g_\alpha(x)) = f_{m-2}(g_\beta(x)) \pmod{q(x)}$. Then, the adversary can compute a collision of size $\mathcal{O}(\max(m, n))$ for the Generalized Tillich-Zémor hash function $\phi$.*

- Even for the simplest equation $f_n(x) = 0 \pmod{q(x)}$, finding a solution for $n$ is not straightforward, since $n$ occurs both as a polynomial term and in the exponent of $2$.

# Condition for Collisions

**Lemma 4**

*Let $\mathbb{F}_p[x]/\langle q(x)\rangle$ be a finite field. If an adversary can find integers $m$ and $n$ such that the following relations hold*

$$\begin{aligned}
f_m(f_n(x)) + f_{m-1}(f_{n-1}(x)) &= 1 \pmod{q(x)} \\
f_m(f_{n-1}(x)) + f_{m-1}(f_{n-2}(x)) &= 0 \pmod{q(x)} \\
f_{m-1}(f_n(x)) + f_{m-2}(f_{n-1}(x)) &= 0 \pmod{q(x)} \\
f_{m-1}(f_{n-1}(x)) + f_{m-2}(f_{n-2}(x)) &= 1 \pmod{q(x)},
\end{aligned}$$

*then $H(0^m 1^n) = H()$ gives a collision with the hash $H()$ of the empty word.*

## Malicious Design for Finite Field

- If $q(x)$ is chosen such that $Y$ has a known and "small enough" multiplicative order $n_y$, then also $A_0$ and $B_0$ have small multiplicative orders which divide $n_y$, and can therefore be calculated easily.

### Proposition 2

- *If one can find $N$ such that $\gcd(f_N(x) - 1, f_{N-1}(x))$ has an irreducible divisor $q(x)$ of degree $d$, one can find a collision of size $\mathcal{O}(N)$ for the hash function $\phi(x)$ over the finite field $\mathbb{F}_p[x]/\langle q(x)\rangle$.*
- *Given a fixed finite field $\mathbb{F}_p[x]/\langle q(x)\rangle$, if one can find an integer $N$ such that $q(x)$ divides $\gcd(f_N(x) - 1, f_{N-1}(x))$ then one can find collisions of size $\mathcal{O}(N)$ for $\phi$.*

## Thank You!

# References I

Petit, Christophe et al. (2009). "Hard and Easy Components of Collision Search in the Zémor-Tillich Hash Function: New Attacks and Reduced Variants with Equivalent Security". In: *Topics in Cryptology – CT-RSA 2009*. Ed. by Marc Fischlin. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 182–194. ISBN: 978-3-642-00862-7.

Zémor, Gilles (1991). "Hash Functions And Graphs With Large Girths". In: *International Conference on the Theory and Application of Cryptographic Techniques*.