Hash function based on the SIS problem

HEBANT Chloé

University of Limoges

Summer 2016

Introduction

Hash function

With a function f which have the properties:

- one-way
- collision-resistant
- compression

Iterating f trying to maintain:

- pre-image resistance
- second pre-image resistance
- collision resistance

Definition

- Pre-image resistance:
  Given $y = H(x)$ it is hard to find x' such that $H(x') = y$
- Second pre-image resistance:
  Given x it is hard to find x' such that $H(x) = H(x')$
- Collision resistance:
  It is hard to find x, x' such that $H(x) = H(x')$

One-way collision-resistant Ajtai function

Let a matrix $A \in \mathbb{Z}_q^{n \times m}$
Let

$$\begin{aligned} f_A : \{0, \pm 1\}^m &\to \mathbb{Z}_q^n \\ z &\mapsto Az \end{aligned}$$

Theorem

f_A is a compression function if $m \geqslant n \log q$

Definition

Definition (SIS problem)

- Given m uniformly random vectors $a_i \in \mathbb{Z}_q^n$
- Find $z \neq 0 \in \{0, \pm 1\}^m$ such that:

$$f_A(z) := Az = \sum_i a_i \cdot z_i = 0 \in \mathbb{Z}_q^n$$

Definition

Definition (SIS problem)

- Given m uniformly random vectors $a_i \in \mathbb{Z}_q^n$
- Find $z \neq 0 \in \{0, \pm 1\}^m$ such that:

$$f_A(z) := Az = \sum_i a_i \cdot z_i = 0 \in \mathbb{Z}_q^n$$

Theorem

Assuming the hardness of the SIS problem, f_A is one-way and collision-resistant

Definition

Definition (SIS problem)

- Given m uniformly random vectors $a_i \in \mathbb{Z}_q^n$
- Find $z \neq 0 \in \{0, \pm 1\}^m$ such that:

$$f_A(z) := Az = \sum_i a_i \cdot z_i = 0 \in \mathbb{Z}_q^n$$

Theorem

Assuming the hardness of the SIS problem, f_A is one-way and collision-resistant

Remark

Thanks to Ajtai and his hardness proof, it's all Minicrypt that we can construct based on the SIS problem.

Some observations

Definition (General SIS problem)

- Given m uniformly random vectors $a_i \in \mathbb{Z}_q^n$
- Find $z \neq 0 \in \mathbb{Z}^m$ of norm $\|z\| \leqslant \beta$ such that:

$$f_A(z) := Az = \sum_i a_i \cdot z_i = 0 \in \mathbb{Z}_q^n$$

Some observations

Definition (General SIS problem)

- Given m uniformly random vectors $a_i \in \mathbb{Z}_q^n$
- Find $z \neq 0 \in \mathbb{Z}^m$ of norm $\|z\| \leqslant \beta$ such that:

$$f_A(z) := Az = \sum_i a_i \cdot z_i = 0 \in \mathbb{Z}_q^n$$

Remark

- Without the constraint on $\|z\|$, it is easy to find a solution: Gaussian elimination
- Must take $\beta < q$:
  otherwise $z = (q, 0, \cdots, 0) \in Z^m$ is a trivial solution

Hermite normal form

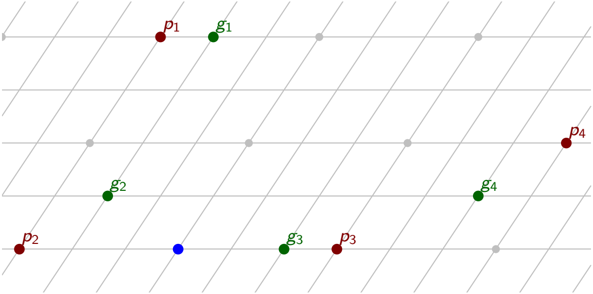
Small but important optimization:

- Decompose $A = [A_1 | A_2]$ where $A_1 \in \mathbb{Z}_q^{n \times n}$ is invertible as a matrix over $\mathbb{Z}_q$.
- Let $B = A_1^{-1} \cdot A = \left[I_n | \bar{A} \right]$ where $\bar{A} = A_1^{-1} \cdot A_2$

Theorem

A and B have exactly the same set of (short) SIS solutions

Reduction: average-case $\rightarrow$ worst-case

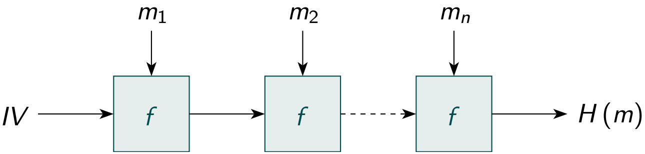


- $p_i \in \mathcal{L}^n$
- $g_i = p_i + e_i \in \mathbb{R}^n$ where $e_i \sim D_s(x) = \left(\frac{1}{s}\right)^n \mathrm{e}^{-\pi \frac{\|x\|^2}{s^2}}$

Merkle-Damgård construction

Definition

Method of building collision-resistant cryptographic hash functions from collision-resistant one-way



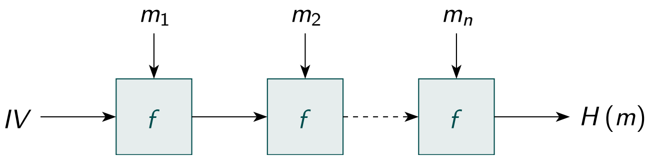
Theorem (Security proof)

Collision in $H \Rightarrow$ collision in f

Merkle-Damgård construction

Definition

Method of building collision-resistant cryptographic hash functions from collision-resistant one-way



Theorem (Security proof)

Collision in $H \Rightarrow$ collision in f

Remark

This is used for MD5, SHA1, SHA2

Several undesirable properties

- **Length extension**
  Given $H(x)$ of an unknown input x,
  it's easy to find the value of $H(\mathrm{pad}(x)||y)$
  $\Rightarrow$ possible to find hashes of inputs related to x even though x remains unknown

Several undesirable properties

- **Length extension**
  Given $H(x)$ of an unknown input x,
  it's easy to find the value of $H(\text{pad}(x)||y)$
  $\Rightarrow$ possible to find hashes of inputs related to x even though x remains unknown

- **Second pre-image**
  Hyp: the security proof also apply to second pre-image attacks
  But: this is not true for long messages

Several undesirable properties (2)

- **Fix-points:** $h = f(h, M)$
- **Multicollisions**: many messages with the same hash
  2004: (Joux) When iterative hash functions are used, finding multicollisions is almost as easy as finding a single collision

Remark

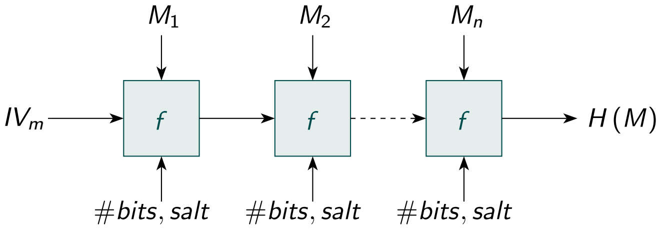
Joux also prove: The concatenation of hash function is as secure against pre-image attacks as the strongest of all the hash functions

HAIFA

HAIFA has attractive properties:

- simplicity
- maintaining the collision resistance of the compression function
- increasing the security against second pre-image attacks
- prevention of esay-to-use fix points of the compression function

HAIFA construction



- $\#bits =$ the number of bits hashed so far
- $IV_m = f(IV, m, 0, 0)$ where m is the hash output size
- Padding scheme: pad a single bit of 1 and as many 0 bits to have the good size. Final length of:
  - M: congruent to $(n - (t + r)) \mod n$
  - length of M: t
  - m: r

HAIFA vs Merkle-Damgård

- $\#$**bits**: prevent the easy exploitation of fix-points

  Even if an attacker finds a fix-point $h = f(h, M, \#bits, salt)$ he cannot concatenate it to itself because $\#bits$ has changed

HAIFA vs Merkle-Damgård

- #**bits**: prevent the easy exploitation of fix-points

  Even if an attacker finds a fix-point $h = f(h, M, \#bits, salt)$ he cannot concatenate it to itself because $\#bits$ has changed

- **salt**:
  - all attacks are on-line $\rightarrow$ no precomputation
  - increasing the security of digital signature

HAIFA vs Merkle-Damgård

- $\#$**bits**: prevent the easy exploitation of fix-points

  Even if an attacker finds a fix-point $h = f(h, M, \#bits, salt)$ he cannot concatenate it to itself because $\#bits$ has changed

- **salt**:
  - all attacks are on-line $\rightarrow$ no precomputation
  - increasing the security of digital signature

- **Multicollisions**: this attacks works against all iterative hashing schemes, independent of their structure

  BUT: an attacker cannot precompute these multicollisions before the choosing of the salt value