Second Preimage Attacks on Dithered Hash Functions

Tomer Ashur

University of Haifa

Authors: Elena Andreeva, Charles Bouillaguet, Pierre-Alain Fouque, Jonathan J. Hoch, John Kelesy, Adi Shamir, and Sebasiten Zimmer.

What is a Hash Function (informal)

- A hash function is a function that maps inputs from arbitrary length to images of fixed length.
- The function should be easy to compute.
- The function should be hard to invert.
- Let H be a cryptographic hash function with output size n. The security requirements are:
  - Preimage resistance: given the output of the function, y, finding a value x such that $H(x) = y$ should at least take 2^n time.
  - Second preimage resistance: given an input x_1, finding another input x_2 such that $H(x_1) = H(x_2)$ should at least take 2^n time.
  - Collision resistance: finding two inputs mapping to the same output should take at least $2^{\frac{n}{2}}$ time.

The Merkle–Damgård Construction

- A compression function is a function that takes input of size $n+k$ an returns an output of size n.
- The Merkle-Damgård construction is a method for constructing hash-function using a collision resistant compression function.
- The Merkle-Damgård algorithm:
  - For a message M and a compression function f, pad and divide the message into r blocks of size n.
  - $h_0 = IV$
  - $h_1 = f(h_0, m_0)$
  - ...
  - $h_i = f(h_{i-1}, m_{i-1})$
  - $H(M) = h_r$
- See drawing on the board.

The Diamond Structure

- The diamond structure is a method to break commitment schemes.
- See drawing on the board.

How to Use a Diamond Structure for Finding Second-Preimage

How to Use a Diamond Structure for Finding Second-Preimage

- Complexity:
  - Building a diamond: $2^{\frac{n}{2}+\frac{l}{2}+2}$
  - Connecting the end of the diamond back to the message: 2^{n-k}
  - Connecting the message to the diamond: 2^{n-l}
- When $l = \frac{n-2}{3}$ The complexity becomes $5 \cdot 2^{\frac{2 \cdot n}{3}} + 2^{n-k}$.
- The attack allows replacing only a small number of message blocks.

Dithered Hash Functions

- Dithered Hashing is a method to add some "freshness" to the Hashing.
- The goal in Dithering is to change a small part of every message thus compressing it differently.
- The Merkle-Damgård will look like this:
    - For a message M and a compression function f, pad and divide the message into r blocks of size n.
    - $h_0 = IV$
    - $h_1 = f(h_0, m_0, d_0)$
    - ...
    - $h_i = f(h_{i-1}, m_{i-1}, d_{i-1})$
    - $H(M) = h_r$
- See drawing on the board.

- Let ω be a word (sequence) over a finite Alphabet A.
- If ω can be written as $\omega = x \cdot y$ ($\cdot$ is the concatenation operation) then x is the prefix of ω and y is a factor of ω.
- A word ω will be called square if it can be expressed as $\omega = x \cdot x$.
- A word ω will be called an abelian-square if it can be expressed as $\omega = x \cdot x'$ where x' is a permutation of x.
- A word is square-free (resp., abelian square-free) if none of its factors is square (resp., abelian square).
- The complexity of a word is the number of its factors of constant size.

Kernen Sequence

- Veikko Keranen showed a construction for an infinite abelian square-free word over a 4-letter alphabet.

Riverst's Proposal

- For a message M having $m_0..m_r$ blocks, generate a word ω of size r.
- Use each letter as the dithering sequence for the compression function.
- This proposal can be implemented with logarithmic space and constant amortized time.
- The message require only 2-bit overhead per message block.
- Alternatively, use a 16-bit dithering sequence to reduce generation time.

Can This be Attacked? Yes!

- See drawing on the board.

Can This be Attacked? Yes!

- It is best to select the most frequent factor of size $l + 1$.
- For an uniform distribution of factors, the complexity of the attack becomes $2^{\frac{n}{2}+\frac{k}{2}+2} + Fact(l+1) \cdot 2^{n-k} + 2^{n-l}$.
- for $l \leq 85; Fact_k(l) \leq 8 \cdot l + 332$. Hence, the complexity is $2^{\frac{n}{2}+\frac{k}{2}+2} + (8 \cdot l + 340) \cdot 2^{n-k} + 2^{n-l}$.
- for $0 \leq l \leq 2^13; Fact_k(l) \leq 8 \cdot l + 32760$. Hence, the complexity is $2^{\frac{n}{2}+\frac{k}{2}+2} + (8 \cdot l + 32768) \cdot 2^{n-k} + 2^{n-l}$.

Can This be Attacked? Yes!

- If n is greater than about $3 \cdot k$; then, the best value for l is $k-3$, and the complexity of the attack becomes approximately $(k+4094) \cdot 2^{n-k+3} \approx 2^{n-k+15}$.

Countermeasures

- The cavity in Rivest's proposal is the relatively small number of factors for the dithering word.
- The dithering word can be extended to i bits hence increasing the number of factors to 2^i providing more security.