Security Proofs for the MD6 Hash Algorithm

Ahmed Ezzat
Outline

- Introduction to hash algorithms
- NIST SHA-3 Competition
- MD6 Algorithm and Mode of Operation
- Research Objective
- Approach
Introduction to hash algorithms

- Hash function definition

Usage Scenarios
- Digital Signature
- Message Integrity
- Password Verification
Introduction to hash algorithms

- Properties
  - Collision resistance
  - First pre-image resistance
  - Second pre-image resistance
  - Pseudo randomness
  - Unpredictability
NIST SHA-3 Competition

- Salted Hashing
- Parellizable

Requirements for a message digest of d-bits:
  - Collision resistance of approximately $d/2$ bits.
  - First - preimage resistance of approximately $d$ bits.
  - Second - preimage resistance of approximately $d - k$ bits for any message shorter than $2^k$ bits.
MD6 Algorithm and Mode of Operation

- Input message structure
MD6 Algorithm and Mode of Operation

- **Input:** \( A[0..88] \) of \( A[0..16r+88] \)
  - for \( i = 89 \) to \( 16r+88 \):
    - \( x = S_i \oplus A[i-17] \oplus A[i-89] \)
    - \( \oplus (A[i-18] \land A[i-21]) \)
    - \( \oplus (A[i-31] \land A[i-67]) \)
    - \( x = x \oplus (x \gg r_i) \)
    - \( A[i] = x \oplus (x \ll l_i) \)
  - return \( A[16r+73..16r+88] \)
MD6 Algorithm and Mode of Operation

- Mode of operation snapshot
MD6 Algorithm and Mode of Operation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Default</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>$-$</td>
<td>the data block portion of a compression function input.</td>
</tr>
<tr>
<td>$b$</td>
<td>64</td>
<td>the number of words in array $B$.</td>
</tr>
<tr>
<td>$C$</td>
<td>$-$</td>
<td>the output of the compression function.</td>
</tr>
<tr>
<td>$c$</td>
<td>$16w$</td>
<td>number of bits in the &quot;chaining variable&quot; $C$.</td>
</tr>
<tr>
<td>$d$</td>
<td>$-$</td>
<td>number of bits in the MD6 final output ($1 \leq d \leq 512$).</td>
</tr>
<tr>
<td>$f$</td>
<td>$-$</td>
<td>the MD6 compression function mapping ${0,1}^{n+k}$ to ${0,1}^{w}$.</td>
</tr>
<tr>
<td>$K$</td>
<td>0</td>
<td>the key variable (an input to $f$).</td>
</tr>
<tr>
<td>$k$</td>
<td>$8w$</td>
<td>number of bits in the key variable $K$.</td>
</tr>
<tr>
<td>$keylen$</td>
<td>0</td>
<td>the length in bytes of the supplied key; $0 \leq keylen \leq kw/8$.</td>
</tr>
<tr>
<td>$\ell$</td>
<td>$-$</td>
<td>the level number of a compression node.</td>
</tr>
<tr>
<td>$L$</td>
<td>31</td>
<td>mode parameter (maximum level number).</td>
</tr>
<tr>
<td>$N$</td>
<td>$-$</td>
<td>the non-key, non-$Q$ piece of the compression function input.</td>
</tr>
<tr>
<td>$n$</td>
<td>$66w$</td>
<td>the size of $N$ (in bits).</td>
</tr>
<tr>
<td>$p$</td>
<td>$-$</td>
<td>the number of padding bits in a data block $B$.</td>
</tr>
<tr>
<td>$Q$</td>
<td>$-$</td>
<td>an approximation to $\sqrt{6}$ (see [32, Appendix A]).</td>
</tr>
<tr>
<td>$q$</td>
<td>15</td>
<td>the length of $Q$ in words.</td>
</tr>
<tr>
<td>$U$</td>
<td>$-$</td>
<td>one-word unique node ID.</td>
</tr>
<tr>
<td>$u$</td>
<td>1</td>
<td>length of $U$ in words.</td>
</tr>
<tr>
<td>$V$</td>
<td>$-$</td>
<td>a control word input to a compression function.</td>
</tr>
<tr>
<td>$v$</td>
<td>1</td>
<td>length of $V$ in words.</td>
</tr>
<tr>
<td>$w$</td>
<td>64</td>
<td>the number of bits in a word.</td>
</tr>
<tr>
<td>$z$</td>
<td>$-$</td>
<td>flag bit in $V$ indicating this is final compression.</td>
</tr>
</tbody>
</table>

The MD6 Mode of Operation

**Input:**

- $M$: A message $M$ of some non-negative length $n$ in bits.
- $d$: The length $d$ (in bits) of the desired hash output, $1 \leq d \leq 512$.
- $K$: An arbitrary $k = 8$ word "key" value, containing a supplied key of $keylen$ bytes.
- $L$: A non-negative mode parameter (maximum level number, or number of parallel passes).
- $r$: A non-negative number of rounds.

**Output:**

- $D$: A $d$-bit hash value $D = M_{d,K,L,r}(M)$.

**Procedure:**

**Initialize:**

- Let $\ell = 0$, $M_0 = M$, and $m_0 = m$.

**Main level-by-level loop:**

- Let $\ell = \ell + 1$.
- If $\ell = L + 1$, return $SEQ(M_{\ell-1}, d, K, L, r)$ as the hash function output.
- Let $M_\ell = PAR(M_{\ell-1}, d, K, L, r, \ell)$. Let $m_\ell$ be the length of $M_\ell$ in bits.
- If $m_\ell = c$ (i.e., if $M_\ell$ is $c$ words long), return the last $d$ bits of $M_\ell$ as the hash function output. Otherwise, return to the top of the main level-by-level loop.
MD6 Algorithm and Mode of Operation

The (Optional) MD6 SEQ Operation

Input:
- $M_L$: A message of some non-negative length $m_L$ in bits.
- $d$: The length $d$ (in bits) of the desired hash output, $1 \leq d \leq 512$.
- $K$: An arbitrary $k = 8$-word key value, containing a supplied key of $k$ bytes.
- $L$: A non-negative mode parameter (maximum tree height).
- $r$: A non-negative number of rounds.

Output:
- $D$: A $d$-bit hash value.

Procedure:

Initialize:
- Let $Q$ denote the array of length $q$ as obtained by the following:
  
  - The MD6 compression function $f_6$ (see Section 2.2)
  
  - The MD6 compression function $f_{6-1}$ (see Section 2.2)

Main loop:
- Let $C_{L-1}$ be the zero vector of length $c = 16$ words. (This is the $\tau^{\ell_0}$)
- Extend input $M_L$ if necessary (and only if necessary) by appending zero bits until its length becomes a positive integral multiple of $(b - c) = 48$ words. Then $M_L$ can be viewed as a sequence $B_0, B_1, \ldots, B_{L-1}$ of $(b - c)$-word blocks, where $j = \max(1, \lfloor mL/(b - c)\rfloor)$.
- For each $(b - c)$-word block $B_i$, $i = 0, 1, \ldots, j - 1$ in sequence, compute $C_i$ as follows:
  - Let $p$ be the number of padding bits in $B_i$, where $0 \leq p \leq 3072$. (We can only be nonzero when $i = j - 1$).
  - Let $k = 1$ if $i = j - 1$, otherwise let $k = 0$. (Set $k = 1$ only for the last block to be compressed in the complete MD6 computation.)
  - Let $V$ be the one-word value $\tau^{\ell_0} + 2^6 - 1$ (see Figure 2-2).
  - Let $U = L - 2^6 + k$ be a "unique" number ID — a one-word value unique to this compression function operation.
  - Let $C_i = f_6(U V B_i)$. (The $C_i$ has length $c = 16$ words.)
- Return the last $d$ bits of $C_{L-1}$ as the hash function output.

The MD6 PAR Operation

Input:
- $M_{L-1}$: A message of some non-negative length $m_{L-1}$ in bits.
- $d$: The length $d$ (in bits) of the desired hash output, $1 \leq d \leq 512$.
- $K$: An arbitrary $k = 8$-word key value, containing a supplied key of $k$ bytes.
- $L$: A non-negative mode parameter (maximum level number, or number of parallel processes).
- $r$: A non-negative number of rounds.
- $\ell$: A non-negative integer level number, $1 \leq \ell \leq L$.

Output:
- $M_L$: A message of length $m_L$ in bits, where $m_L = 1024 - \max(1, \lfloor m_{L-1}/0000\rfloor)$.

Procedure:

Initialize:
- Let $Q$ denote the array of length $q$ as obtained by the following:
  
  - The MD6 compression function $f_6$ (see Section 2.2)
  
  - The MD6 compression function $f_{6-1}$ (see Section 2.2)

Shrink:
- Extend input $M_{L-1}$ if necessary (and only if necessary) by appending zero bits until its length becomes a positive integral multiple of $b = 64$ words. Then $M_{L-1}$ can be viewed as a sequence $B_0, B_1, \ldots, B_{L-1}$ of $b$-word blocks, where $j = \max(1, \lfloor mL/b \rfloor)$.
- For each $b$-word block $B_i$, $i = 0, 1, \ldots, j - 1$ in sequence, compute $C_i$ as follows:
  - Let $d$ denote the number of padding bits in $B_i$, where $0 \leq d \leq 4096$. (We can only be nonzero for $i = j - 1$).
  - Let $k = 1$ if $i = j - 1$, otherwise let $k = 0$. (The $k - 1$ only for the last block to be compressed in the complete MD6 computation.)
  - Let $V$ be the one-word value $\tau^{\ell_i} + 2^6 - 1$ (see Figure 2-2).
  - Let $U = L - 2^6 + k$ be a "unique" number ID — a one-word value unique to this compression function operation.
  - Let $C_i = f_6(U V B_i)$. (The $C_i$ has length $c = 16$ words.)
- Return $M_L = C_0 \mid C_1 \mid \ldots \mid C_{L-1}$. 

Properties Remaining

- Second pre-image resistance
- Unpredictability
Research Objective

The continuation of the security proofs for the MD6 hash function mode of operation
Approach

- Mathematical
- Empirical
References

