Cryptanalysis on HMAC-MD5, MD5-MAC and Sandwich-MAC-MD5

Yu Sasaki\textsuperscript{1} and Lei Wang\textsuperscript{2}

\textsuperscript{1}NTT Secure Platform Laboratories
\textsuperscript{2}Nanyang Technological University, Singapore
ASK 2013 (27/August/2013@Weihai)
Hash Function Based MAC

- Message Authentication Codes (MAC) provide the integrity and authenticity.

secret key: $K$

message: $M$

Tag: $\text{Hash}(M,K)$

secret key: $K$

Check the match of the tag
Classical MAC Constructions

- **Prefix**
  - Length extension attack

- **Suffix**
  - Collision attack

- **Hybrid**
  - Secure !!
HMAC

• The most widely used hash-based MAC
  – Requires 2 keys for inner and outer functions
  – Requires 2 hash function calls
  – 3 additional blocks for converting hash into MAC; non-negligible overhead for short messages
Sandwich-MAC

- Several MACs improve HMAC
- Sandwich-MAC [Yasuda ACISP 2007] has advantages on performance.
  - Requires 1 key
  - Requires 1 hash function call
  - 2 additional blocks for converting hash into MAC; small overhead, suitable for short messages
Motivation

• HMAC and Sandwich-MAC have the same provable security: secure PRF up to $O(2^{n/2})$.

• Need more comparison

• We investigate attacks when a weak hash function (MD5) is instantiated.

• Then, extract features which can be applied in generic.
Our Contributions

1. Improve the internal state recovery attack on HMAC-MD5 both in adaptive and non-adaptive settings.

2. By using the above, propose a key-recovery attack on Sandwich-MAC-MD5.
   – First key recovery attack on hybrid-type MACs
   – conditional key-dependent distribution technique

3. Improve the attack on MD5-MAC$_{K_0,K_1,K_2}$.
   – Improve the complexity to recover $K_1$.
   – Propose the first key recovery attack for $K_2$. 
## Attack Results

<table>
<thead>
<tr>
<th>Target</th>
<th>Model</th>
<th>Attack goal</th>
<th>Data</th>
<th>Time</th>
<th>Memory</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMAC-MD5</td>
<td>Adaptive</td>
<td>Dist-H/ISR</td>
<td>$2^{97}$</td>
<td>$2^{97}$</td>
<td>$2^{89}$</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>Dist-H/ISR</td>
<td>$2^{89.09}$</td>
<td>$2^{89}$</td>
<td>$2^{89}$</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td>Non-adaptive</td>
<td>Dist-H/ISR</td>
<td>$2^{113}$</td>
<td>$2^{113}$</td>
<td>$2^{66}$</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td>Non-adaptive</td>
<td>Dist-H/ISR</td>
<td>$2^{113-x}$</td>
<td>$2^{113-x}$</td>
<td>$2^{66+x}$</td>
<td>Ours</td>
</tr>
<tr>
<td>MD5-MAC</td>
<td>$K_1$-recovery</td>
<td></td>
<td>$2^{97}$</td>
<td>$2^{97}$</td>
<td>$2^{89}$</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td>$K_1$-recovery</td>
<td></td>
<td>$2^{89.09}$</td>
<td>$2^{89}$</td>
<td>$2^{89}$</td>
<td>Ours</td>
</tr>
<tr>
<td></td>
<td>$(K_1, K_2)$-recovery</td>
<td></td>
<td>$2^{89.04}$</td>
<td>$2^{89}$</td>
<td>$2^{89}$</td>
<td>Ours</td>
</tr>
<tr>
<td>Sandwich-MAC-MD5</td>
<td>Basic</td>
<td>Key recovery</td>
<td>$2^{89.04}$</td>
<td>$2^{89}$</td>
<td>$2^{89}$</td>
<td>Ours</td>
</tr>
<tr>
<td>Variant B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Improved Single-key Attacks against HMAC-MD5
MD5

- Widely known to be broken but still widely used

Merkle-Damgård structure

\( (H_0) \) \( M_0 \) \( 512 \) \( IV \) \( h \) \( h \) \( h \) \( h \) \( H_1 \) \( H_2 \) \( \cdots \) \( H_{\ell-1} \) \( M_{\ell-1} \) \( \text{pad} \) \( \rightarrow \) \( \text{Hash}(M) \)

Compression function \( h \)

\( (m_0, m_1, \ldots, m_{15}) \leftarrow M_{i-1} \)

\( H_{i-1} \) \( \rightarrow \) \( \text{Step 1} \) \( \text{Step 2} \) \( \text{Step 3} \) \( \text{Step 4} \) \( \cdots \) \( \text{Step 15} \) \( \text{Step 16} \) \( \cdots \) \( \text{Step 64} \) \( \downarrow \) \( \downarrow \) \( \downarrow \) \( \downarrow \) \( \downarrow \) \( \downarrow \) \( \downarrow \) \( \rightarrow \) \( H_i \)
dBB-collision

- The compression function $h$ generates a collision with probability $2^{-48}$ for $(H_{i-1}, M_{i-1})$ and $(H_{i-1}', M_{i-1})$ when $H_{i-1} \oplus H_{i-1}'$ has a special difference called $\Delta^{\text{MSB}}$.

- In the dBB-collision, each of the first 16 steps has the differential characteristic with $Pr.=2^{-1}$. 

\[
\begin{array}{c|c|c|c|c}
\Delta^{\text{MSB}} & \begin{array}{c} m_0 \downarrow \downarrow \downarrow \downarrow \end{array} & \begin{array}{c} m_1 \downarrow \downarrow \downarrow \downarrow \end{array} & \begin{array}{c} m_2 \downarrow \downarrow \downarrow \downarrow \end{array} & \begin{array}{c} m_3 \downarrow \downarrow \downarrow \downarrow \end{array} \\
H_{i-1} & \text{Step 1} & \text{Step 2} & \text{Step 3} & \text{Step 4} \\
\hline
\begin{array}{c} m_{14} \downarrow \downarrow \downarrow \downarrow \end{array} & \begin{array}{c} m_{15} \downarrow \downarrow \downarrow \downarrow \end{array} & \text{Step 15} & \text{Step 16} \\
\hline
\text{Step 64} & \Delta=0 & H_i
\end{array}
\]
Previous Attack against HMAC-MD5

1. Generate $2^{128} \times 2^{48} = 2^{176}$ pairs by changing $M_0$.
   - One pair satisfies the dBB-collision.
   - We have other $2^{176-128} = 2^{48}$ collisions. (noise)

2. For each $2^{48}$ collisions, change $M_1$ $2^{48}$ times.
   - If another collision is found, it is a dBB-collision.

Birthday attack to generate $\Delta_{MSB} (2^{-128})$

Follow the dBB-collision $(2^{-48})$
Improving ISR against HMAC-MD5

Previous work: retake all messages $\Rightarrow$ $Pr = 2^{-48}$.

Ours: Reuse the messages for the first 14 steps so that the characteristic remains satisfied. $\Rightarrow$ $Pr = 2^{-34}$. 
Key Recovery Attacks against Sandwich-MAC-MD5
Phase 1: Internal State Recovery

- Recover the internal state value $H_2$, similarly with the internal state recovery on HMAC-MD5.
Phase 2: IV Bridge

- From the recovered $H_2$, find $(M_1, M_1')$ which generates $\Delta^{MSB}$ at $H_3$.
- This can be done by a variant of collision attack called IV Bridge with a complexity of $2^{10}$ [Tao+ ePrint].

![Diagram showing the IV Bridge process with IV, $K||ipad1$, $M_0$, $M_1||pad2$, $K||pad3$, and the hash function $h$ with $H_1$, $H_2$, $H_3$, and $\Delta^{MSB}$]
Phase 3: Collecting dBB-near-collisions

- By querying $2^{48}$ IV bridges, one tag collision is obtained. To be precise, $2^{47}$ IV bridges to obtain dBB-near-collisions enough.
- For the dBB-near-collision, 1 bit of internal state is recovered because the characteristic is satisfied.
Key Recovery with Conditional Key Distributions

• Due to the structure of the MD5 compression function, 32 bits of the tag $\tau$ are computed by (internal state $Q$) ⊕ (a part of secret key $k$)

\[ Q \quad \text{1 bit (MSB) is known} \]

\[ \text{secret} \quad k \quad \text{32 bit} \]

\[ \text{32 bit} \quad \downarrow \]

\[ \text{32 bit} \]

\[ \tau \quad \text{known} \]

• By collecting $2^{32}$ pairs of such $(Q, \tau)$, the secret key $k$ can be recovered.
Conditional Key-dependent Distributions

• Collect pairs in which the 30\textsuperscript{th} bit of $\tau$ is 0.
  
  1. If the 30\textsuperscript{th} bit of $k$ is 0: two possible carry patterns
  
  2. If the 30\textsuperscript{th} bit of $k$ is 1: one possible carry pattern

• Behavior of the addition depends on the key value. This eventually reveals the 30\textsuperscript{th} and 31\textsuperscript{st} bits of $k$. 

$$
\begin{align*}
\begin{array}{cccccccc}
\text{0/1} & & & & & & & \\
\text{MSB} & & & & & & & \\
+ & 0 & & & & & & \\
\hline
\text{31} & 0 & 29 & 28 & & & & \\
\end{array}
\end{align*}
$$

$$
\begin{align*}
\begin{array}{cccccccc}
1 & & & & & & & \\
\text{MSB} & & & & & & & \\
+ & 1 & & & & & & \\
\hline
\text{31} & 0 & 29 & 28 & & & & \\
\end{array}
\end{align*}
$$
Phase 4: Rest of Attacks

• The key for the last step is recovered by using the conditional key distribution.
• Then, all keys are recovered step by step for the last 16 steps.
Key Recovery Attacks against MD5-MAC
Structure of MD5-MAC

• Generate \((K_0, K_1, K_2)\) from the master key \(K\).
  
  – \(K_0\): used for the initialization.
  
  – \(K_1\): used for the message processing. The round constant of MD5 is replaced with \(K_1\).
  
  – \(K_2\): used for the finalization.
Improvement for MD5-MAC

• The previous work recovered $K_1$ based on the internal state recovery attack on HMAC-MD5.

• We improve the $K_1$ recovery similarly to HMAC-MD5.

• We recover $K_2$ similarly to Sandwich-MAC-MD5.
Discussion:
HMAC v.s. Sandwich-MAC
Comparison of HMAC and Sandwich-MAC

- Sandwich-MAC: A differential characteristic to recover the internal state is reused to recover $K$.
- HMAC: Two good characteristics are needed to recover $K$.

Message processing part is identical. Finalization is different.
Comparison for Block-cipher Based Hash

- In hybrid MACs, the MMO mode is the only choice for the finalization computation to resist side-channel analysis [Okeya ACISP 2006].
- Most of the currently used hash function adopts the Davies-Meyer mode.
- The HMAC construction is the most reasonable!!
Concluding Remarks

Attacks with MD5
• Improved internal state recovery attack on HMAC-MD5 in adaptive and non-adaptive settings.
• Key-recovery attack on Sandwich-MAC-MD5 with conditional key distribution techniques.
• Improve the attack on MD5-MAC.

Comparison with HMAC and Sandwich-MAC
• A certain type of differential characteristic can recover the key for Sandwich-MAC.
• From various viewpoints, HMAC is a solid design.
Thank you for your attention!!