Message Authentication

15-859I
Spring 2003

Motivation

- Suppose Alice is an ATM and Bob is a Bank, and Alice sends Bob messages about transactions over a public channel.
- Bob would like to know that when he receives a message saying “credit $128 to Carol’s Account – Alice”, it originates from the ATM. Bob is concerned with the 
  authenticity 
  of the message.
- He also wants to know that Carol has not modified the message from “credit $16 to Carol’s Account – Alice.” This concerns the 
  integrity 
  of the message.

Authentication and Encryption

- Should we expect to get good message authentication via encryption? i.e., is it enough to guarantee authenticity of M by transmitting E_K(M)?
- No! e.g., if E_K is CTR from lecture 6, then it is easy for Carol to change E_K(16) to E_K(128) via E_K(128) = E_K(16) \oplus 144.
- In general, good encryption does not necessarily imply integrity.

Message Authentication Codes (MACs)

- Formally: A MAC is a trio of algorithms (G,T,V) such that:
  - G(1^k) generates a k-bit key K.
  - T_K(M) generates a L(k)-bit tag \sigma
  - V_K(M,\sigma) verifies the tag \sigma for the message M.
- Require that for all K, M, random choices or states of T, V_K(M,T_K(M)) = 1.
  - If T_K is deterministic and stateless, V_K is trivial.

Security of MACs

- The adversary’s goal might be to sign some specific message m.
- We want it to be hard to produce any (M,\sigma) pair such that V_K(M,\sigma)=1.
- This should be true even if the adversary has seen several M',T_K(M') pairs
- Should be conservative: Allow the adversary to choose the M' messsages.

Existential Unforgeability

- Notation: let Q(A^O) denote the list of oracle queries that A makes with O as its oracle.
- Define the chosen-message attack (cma) advantage of A against MAC = (G,T,V) by:
  \[ \text{Adv}_{A,MAC}^{\text{cma}}(k) = \Pr[V^e_A(M,\sigma)=1 \land M \not\in L] \]
  \[ = G(1^k), (M,\sigma) \leftarrow A^e(1^k), L = Q(A^e(1^k)) \] 
- Say that MAC is existentially unforgeable under chosen message attack if every ppt A has negligible advantage.
MAC Insecurity

For a fixed security parameter $k$, define the Insecurity of MAC=$\langle G, T, V \rangle$ against time-$t$ adversaries which make $q$ queries with total message length $l$ by:

$$\text{InSec}_{\text{MAC}}^{\text{uf-cma}}(t, q, l) = \max_{A(k, q, l)} \left\{ \text{Adv}_{\text{MAC}}^\text{cma}(k) \right\}$$

PRFs are good MACs

Let $F : K \times \{0,1\}^d \rightarrow \{0,1\}^s$ be a function family. Then if $F_k$ is pseudorandom, $F_k$ is a good MAC for the message space $\{0,1\}^d$:

$$\text{InSec}_{\text{PRF}}^{\text{uf-cma}}(t, q, dq) \leq \text{InSec}_{\text{PRF}}^{\text{uf-cma}}(t', q) + 2^{-s}$$

Proof: Let $A$ be a chosen-message forger for $F$ as a MAC. We show how to construct a PRF distinguisher $D$ for $F$ that has almost the same advantage as $A$, and runs in the same time.

PRFs are good MACs

$$D(1^k)$$:
Run $A(1^k)$: respond to $q$ with $f(q)$, add $q$ to $Q$
Set $(M, \sigma) =$ output of $A$.
If $f(M) = \sigma$ and $M \notin Q$, return 1, else return 0.

Notice: $\Pr[D_f(1^k) = 1] = \text{Adv}_{A,F}^{\text{cma}}(k)$. And since $M$ was never queried, $\Pr[D_f(1^k) = 1] \leq 1/2^s$.
So $\text{Adv}_{A,F}^{\text{prf}}(k) \geq \text{Adv}_{A,F}^{\text{cma}}(k) - 2^{-s}$

Almost-XOR-Universal$_2$ (AXU$_2$) Hash Functions

Let $H : K \times D \rightarrow \{0,1\}^l$ be a family of functions. Define the XOR-2-Universality of $H$ by

$$\text{Adv}_{\text{XUH}}(H) = \max_{a_1, a_2 \in \{0,1\}^d} \left\{ \Pr_{H, a_1, a_2} [H_k(a_1) \oplus H_k(a_2) = b] \right\}$$

We say $H$ is $\varepsilon$-almost-XOR-Universal$_2$ ($\varepsilon$-AXU$_2$) if $\text{Adv}_{\text{XUH}}(H) \leq \varepsilon$.

H is XOR-Universal$_2$ if it is $1/2^l$-AXU$_2$.

Notice that Pairwise-independent hash functions are XOR-Universal$_2$.

$\varepsilon$-AXU$_2$ Hash Families for large domains

Let $h : K \times \{0,1\}^{2L} \rightarrow \{0,1\}^L$ be $\varepsilon$-AXU$_2$. Define the family $H : K^n \times D \rightarrow \{0,1\}^L$ as follows:

$H_{K_1, \ldots, K_n}(a_1, \ldots, a_2^n) = h_{K_1, \ldots, K_n}(h_{K_1}(a_1, a_2), h_{K_1}(a_3, a_4), \ldots, h_{K_1}(a_{2^n-1}, a_{2^n}))$;

$H_k(a_1, a_2) = h_k(a_1, a_2)$

Claim: $H$ is $(n\varepsilon)$-AXU$_2$.

Proof: If the inputs to $h_{K_0}$ are not the same, then the xor-probability is at most $\varepsilon$. The probability that the inputs to $h_{K_0}$ are the same is at most $\varepsilon$, given that not all inputs to $h_{K_0}$ are the same, and so on.

AXU$_2$ - MAC

Let $F : K \times \{0,1\}^l \rightarrow \{0,1\}^l$ be a PRF. Let $H : K \times D \rightarrow \{0,1\}^l$ be $\varepsilon$-AXU$_2$. Define the MAC $C_{\text{UHMAC}}$ as follows:

$G$: Select $K \leftarrow K_0 \leftarrow K$. Return $(K, \kappa)$

$T_{(K, \kappa)}(M) =$
- Let $x = H_1(M)$;
- $\tau = F_k(\text{ctr}) \oplus \kappa$;
- Set $\text{ctr} = \text{ctr} + 1$
- Return $\sigma$

$V_{(K, \kappa)}(M, (s, \tau)) = 1$ iff $F_k(s) \oplus \tau = H_k(M)$. 

C-UHM theorem

- Theorem: for any $q \leq 2^l$,
  \[
  \text{InSec}_{\text{C-UHM}}^{\text{ud-cma}}(t,q,l) \leq \varepsilon + \text{InSec}_{\text{F}}^{\text{prf}}(t',q+1) + q(q-1)/2^{l+1}
  \]
- Proof: Consider the same experiment as before. Clearly when there are no collisions in the values $s_1, ..., s_q$, the same argument upper bounds the success probability of $A$. If there is a collision, the success probability of $A$ is at most 1. The probability of a collision is $q(q-1)/2^{l+1}$.

C-UHM Theorem, continued.

Let $\text{NEW}$ be the event that $s > q-1$, that is, the $s$ returned by $A$ was not a value input to $f$ in C-UHM. Let $\text{OLD}$ be the event $s < q$.

Claim 1: \[
\Pr[V_\kappa(f(M,(s,\tau))) = 1 \mid \text{OLD}] \leq 2^{-l}.
\]
Proof: \[
\Pr[V_\kappa(f(M,(s,\tau))) = 1] = \Pr[H_\kappa(M) \oplus \tau = f(s)] = 2^{-l}.
\]

C-UHM Theorem, continued.

Thus:
\[
\Pr[V_\kappa(f(M,(s,\tau))) = 1] = \Pr[V_\kappa(f(M,(s,\tau))) = 1 \mid \text{OLD}] \Pr[\text{OLD}] + \Pr[V_\kappa(f(M,(s,\tau))) = 1 \mid \text{NEW}] (1 - \Pr[\text{OLD}]) \leq \varepsilon + 2^{-l}(1-q) \leq \varepsilon + \varepsilon(1-q) = \varepsilon.
\]

The theorem follows, since we can distinguish $F_k$ from $f$ by trying to use $A$ to forge a MAC and then checking if $A$ was successful.

R-UHM

- Let $F : \{0,1\}^k \times \{0,1\}^l \rightarrow \{0,1\}^l$ be a PRF. Let $H : K \times \{0,1\}^k \rightarrow \{0,1\}^l$ be $\varepsilon$-AXU$_2$. Define the MAC $R$-UHM as follows:
  - $G$: Select $K \leftarrow \mathcal{K}$, $\kappa \leftarrow K$. Return $(K, \kappa)$
  - $T(K, \kappa)(M) = \langle s \leftarrow \{0,1\}^l \rangle$
  - For $i = 1, ..., m$
    - $y_i = F_K(M \oplus y_{i-1})$
  - Return $y_m$
  - $V(K, \kappa)(M, (s, \tau)) = 1$ iff $F_K(s) \oplus \tau = H_{\kappa}(M)$.

R-UHM theorem

- Theorem: for any $q \leq 2^l$,
  \[
  \text{InSec}_{\text{R-UHM}}^{\text{ud-cma}}(t,q,l) \leq \varepsilon + \text{InSec}_{\text{F}}^{\text{prf}}(t',q+1) + q(q-1)/2^{l+1}
  \]
- Proof: Consider the same experiment as before. Clearly when there are no collisions in the values $s_1, ..., s_q$, the same argument upper bounds the success probability of $A$. If there is a collision, the success probability of $A$ is at most 1. The probability of a collision is $q(q-1)/2^{l+1}$.

CBC-MAC

- Let $F : \{0,1\}^k \times \{0,1\}^l \rightarrow \{0,1\}^l$ be a PRF. Define the MAC $F^{(m)}$ on $m$-bit messages as follows:
  - $T_F(x_1, ..., x_m) =$
    - Let $y_0 = 0$
    - For $i = 1, ..., m$
      - $y_i = F_K(M \oplus y_{i-1})$
    - return $y_m$
  - Theorem:
    - $\text{InSec}_{F^{(m)}}(t,q,l) \leq \text{InSec}_{F}^{prf}(t+O(qm),qm) + 3q^3m^2/2^{l+1}$.
    - So $F^{(m)}$ is a secure MAC if $F$ is a secure PRF.
Lemma 1: If $f \leftarrow F_{l,l}$ then
\[ \text{Insec}_{f_{\text{prf}}}^{q}(q) \leq \frac{3m^{2}q^{2}}{2l+1} \]

Consider the $2l$-ary tree of depth $m$. A sequence of strings $X = (x_{1}, \ldots, x_{n}) \in \{0,1\}^{nl}$ uniquely specifies a node in this tree.

Let $f : \{0,1\}^{l} \rightarrow \{0,1\}^{l}$, denote the labeling of a sequence $x_{1} \ldots x_{n}$ by $Z_{f}(X) = 0^{l}$, $Y_{f}(x_{1}, \ldots, x_{n}) = x_{n} \oplus Z_{f}(x_{1}, \ldots, x_{n-1})$, $Z_{f}(X) = f(Y_{f}(X))$.

Call $(X_{1}, \ldots, X_{n})$ a query sequence if every $X_{i}$ has parent either the root or $X_{j}$ for some $j < i$.

CBC-MAC Proof

Lemma 2: Let $Z_{n}^{1}$ and $Z_{n}^{2}$ be collision-free output labelings consistent with a depth-$m$ labeling $Z_{n}$. Then:
\[ \Pr[\text{Z}_{n} = Z_{n}^{i} | V_{n} = (X_{1}, \ldots, X_{n} ; Z_{n})] = \Pr[\text{Z}_{n-1} = Z_{n-1}^{i} | V_{n-1} = (X_{1}, \ldots, X_{n-1} ; Z_{n-1})] \Pr[\text{Z}_{n}(X_{n}) = Z_{n}^{i}(X_{n}) | Z_{n-1} = Z_{n-1}^{i}, V_{n-1}] \]

Proof: By induction. Obviously true for $n=1$, since the first node in a query tree is not at depth $m$.

Lemma 2 proof, con't

Two cases for $n>1$:

- $X_{n}$ at depth $< m$. Then $\Pr[\text{Z}_{n} = Z_{n}^{i} | V_{n} = (X_{1}, \ldots, X_{n} ; Z_{n})] = \Pr[\text{Z}_{n-1} = Z_{n-1}^{i} | V_{n-1} = (X_{1}, \ldots, X_{n-1} ; Z_{n-1})]$.
- $X_{n}$ at depth $m$. Then $\Pr[\text{Z}_{n} = Z_{n}^{i} | V_{n} = (X_{1}, \ldots, X_{n} ; Z_{n})] = \Pr[\text{Z}_{n}(X_{n}) = Z_{n}^{i}(X_{n}) | Z_{n-1} = Z_{n-1}^{i}, V_{n-1} = (X_{1}, \ldots, X_{n-1} ; Z_{n-1})] \Pr[\text{Z}_{n-1}(X_{n}) = Z_{n-1}^{i} | V_{n-1} = (X_{1}, \ldots, X_{n-1} ; Z_{n-1})]$. These are equal for $i=1,2$ by IH.

Lemma 3

Lemma 3: Let $CF(Z)$ denote the event that $Z$ is collision-free. Let $Pr_{\text{CF}}[E]$ denote the quantity $Pr[E | V_{n} = (X_{1}, \ldots, X_{n} ; Z_{n})]$. Let $n/2 < l$ and
\[ \Pr[\text{Z}_{n} = Z_{n}^{i} | V_{n} = (X_{1}, \ldots, X_{n} ; Z_{n})] \leq 2^{-l} \]

(1) For any $(x_{1}, \ldots, x_{n}) \in S$, any $y^{*} \in \{0,1\}^{l}$:
\[ \Pr[y^{*}(x_{1}, \ldots, x_{n}) = y^{*} | Z_{n} = z_{n}] \leq 2^{-l} \]

(2) For any $z^{*} \in \{0,1\}^{l}$:
\[ \Pr[z^{*}(x_{1}, \ldots, x_{n}) = z^{*} | Z_{n} = z_{n}] \leq 2^{l} \]

Proof of Lemma 3(1)

Let $y \in \{0,1\}^{l}$ be some fixed string. Define the labeling $Y_{y}(X) = y_{X}$ if $X \neq x_{1} \ldots x_{n}$ and $y_{X}^{*} = y_{X}$ otherwise. Let $Y_{X}$ be the labeling induced by $Z_{n}$:
\[ Y_{X}(x_{1}, \ldots, x_{n}) = y_{X} = y_{X_{1}, \ldots, x_{n}} \]
\[ y_{X}^{*} x_{j}^{*} = y_{X_{1}, \ldots, x_{n}}^{*} = y_{X_{1}, \ldots, x_{n}}^{*} \]
\[ (y_{X}(x_{1}, \ldots, x_{n}) = y_{X_{1}, \ldots, x_{n}}^{*}) \text{ if } X \neq x_{1} \ldots x_{n} \]
\[ (y_{X}(x_{1}, \ldots, x_{n}) = y_{X_{1}, \ldots, x_{n}}^{*}) \text{ if } X = x_{1} \ldots x_{n} \]

Let $\gamma_{z_{n}}$ be the set of all strings $y$ such that $Z_{z_{n}}$ is collision-free. $y \in \gamma_{z_{n}}$ iff either:
\[ y^{*}(x_{j}) = y_{X}^{*} x_{j}^{*} \text{ if } X = x_{1} \ldots x_{n} \text{ and } 0 < j < n+1 \]
\[ y^{*}(x_{j}) = y_{X}^{*} x_{j}^{*} \text{ if } X \neq x_{1} \ldots x_{n} \text{ and } 0 < j < n+1 \]

Thus $|\{0,1\}^{l} \setminus \gamma_{z_{n}}| \leq (n-1) + (n-s)(s) \leq n-1 + n^{2}/4 \leq 2^{l}/2$. This proves (1).
Proof of Lemma 3(2)

Let \( z \in \{0,1\}^l \) be some fixed string. Define the labeling
\[ Z_z(X_j) = \begin{cases} zS(X_j) & \text{if } X_j \neq x_1\ldots x_i, \\ z & \text{otherwise.} \end{cases} \]

Let \( Y_z \) be the labeling induced by \( Z_z \):
\[ Y_z(X_j) = \begin{cases} yS(X_j) & \text{if } X_j \notin \text{children}(x_1\ldots x_i) \\ z \oplus x_{i+1} & \text{if } X_j = x_1\ldots x_ix_{i+1}. \end{cases} \]

Let \( Z(zS) \) be the set of all strings \( z \) such that \( Z_z \) is collision-free.

\[ z \notin Z(zS) \iff \begin{cases} z \in \{zS(X_j) : 0 < j < n+1 \text{ and } X_j \neq (x_1\ldots x_i)\}; \\ \text{or} \\ z \oplus x_{i+1} \in \{yS(X_j), X_j \notin \text{children}(x_1\ldots x_i), \text{ and } 0 < j < n+1\}. \end{cases} \]

Thus \( \{0,1\}^l \setminus Y(zS) \leq (n-1) + (n-s)(s) \leq n-1 + n^2/4 \leq 2l/2. \)

This proves (2).

Lemma 4: \( \Pr[\text{not CF}(Z)] \)

Let \( n^2/4 + n-1 < 2l/2 \). Let \( X_1\ldots X_n \) be a query sequence and \( z \) be the labeling of depth-m nodes. Then
\[ \Pr[\text{not CF}(Z_{n+1}) | V_n, \text{CF}(Z_n)] \leq 3n^{2-l}. \]

Proof: Denote \( \Pr[E | V_n, \text{CF}(Z_n)] \) by \( \Pr[E] \).

Case 1: \( X_{n+1} \) is at depth 1. Then let \( X_n^{\text{new}} = x_i^* \). \( Y(X_n^{\text{new}}) = y_i^* \) by definition. Now for each 1st \( x_i \),
\[ \Pr[Y(X_i) = x_i^*] \leq 2^{\text{depth}(x_i)}. \]

This is because if \( x_i \) is at level 1, \( \Pr[Y(X_i) = x_i^*] = 0. \) Otherwise \( x_i \) is at depth at least 2, and is the child of some \( (x_1\ldots x_i) \) and so the equation follows because of lemma 3.

Then \( \Pr[\text{not CF}(Z_i)] \leq \Pr[x_i^* \in \{Y(X_1),\ldots,Y(X_i)\}] + \Pr[Z_i(x_1\ldots x_i) = Y_i(x_1\ldots x_i)] + x_i^* \notin \{Y_i(x_1),\ldots,Y(X_i)\}] \leq 2n^{2-l} + n^{2-l} = 3n^{2-l}. \)

Lemma 4: Case 2

Case 2: \( X_{n+1} \) is the child of some \( x_1\ldots x_i \).

\[ \Pr[Y_n(x_1\ldots x_i) = Y_n(X_i)] \leq 2^{\text{depth}(x_i)}. \]

Since if \( x_1\ldots x_i \) and \( X_i \) are siblings, the probability is 0, and otherwise any collision free labeling \( z \) determines \( Y_n(X_i) \), thus
\[ \Pr[Y_n(x_1\ldots x_i) = Y_n(X_i)] = \sum z \Pr[Z_i(x_1\ldots x_i) = z] \Pr(Z_i(x_1\ldots x_i) = z) \leq 2^{\text{depth}(x_i)}. \]

This gives us that
\[ \Pr[\text{not CF}(Z_i)] \leq \Pr[Y_i(x_1\ldots x_i) = Y_i(X_i)] + \Pr[Z_i(x_1\ldots x_i) = Z_i(X_i)] \Pr[Y_i(x_1\ldots x_i) = Y_i(X_i)] \leq 2n^{2-l} + n^{2-l} = 3n^{2-l}. \]

\[ \Pr[\text{CF}(Z)] \]

So
\[ \Pr[\text{not CF}(Z)] \leq \sum \Pr[\text{not CF}(Z_i) | \text{CF}(Z_{i-1})] \leq 3/2^{l} (qm)(qm-1)/2 \leq 3/2 q^2m^2/2. \]