Symmetric Cryptography

15-859I
Spring 2003

Cast of Characters:

Alice, Bob
M, a message
K, a key
Eve

Introduction

- Alice wants to send M to Bob
- Eve wants to find out what M is
- Alice and Bob don’t want her to.
- Previously, Alice and Bob chose K (together) randomly, so that no one else would know it.
- Can they use one secret (K) to keep another secret (M)?

Encryption Schemes

- Alice and Bob want an Encryption Scheme:
  - An encryption scheme is a triple $SE = (G,E,D)$ of Algorithms:
    - $G(1^k)$: generates a key of length $k$
    - $E_K: P \rightarrow C$ maps an input message space (plaintexts) to an output message space (ciphertexts)
    - $D_K: C \rightarrow P$ maps an ciphertexts to plaintexts
  - For all $K$, for all $M \in P$, we require that $D_K(E_K(M)) = M$.

Security of Encryption schemes

- What does it mean for $SE$ to be secure?
- Of course, given $E_K(M)$, Eve should not be able to guess M.
- We will call an attack where Eve recovers M from only $E_K(M)$ a plaintext recovery (pr) attack.
- What if M comes from very small subset of P?
- Ideally, we would like Eve to “get no information about M from $E_K(M)$.”

This problem is solved unconditionally

- Let $P = \{0,1\}^k$, define $OTP = (G,E,D)$ as follows:
  - $G(1^k) = \text{return } K \leftarrow U_k$.
  - $E_K(M) = K \oplus M$
  - $D_K(C) = K \oplus C$
  - It is not hard to see that for $M$ chosen from any distribution on $P$,
  - $H(M|E_K(M)) = H(M)$
  - i.e., $E_K(M)$ gives no information about M.
**Problem**
- We can only use K once, to encrypt $|K|$ bits.
- This means we have to know, beforehand, how many bits we plan to exchange (or an upper bound)
- Then we have to generate that many bits and keep them all secret.
- If we are never in a secure location again, we can never extend the number of bits we can transmit.

**Solution**
- Instead of considering arbitrarily powerful Eve, we constrain Eve to run in polynomial time.
- This suggests that pseudorandomness may be useful.
- What should it mean for a polytime Eve to learn no information from $E_K(M)$?

---

**Security against Plaintext Recovery**
- Suppose Eve plays the following game:
  - $\text{Exp}^\text{p}(\text{Eve}) =$
    - Choose $K \leftarrow U_k$
    - Choose $M \leftarrow U_m$
    - If $E_K(M) = M$ output 1 else output 0
  - Define $\text{Adv}^\text{p}(\text{Eve}) = \Pr[\text{Exp}^\text{p}(\text{Eve}) = 1]$
  - Define $\text{Insec}^\text{p}(SE, t, q, l) = \max_{\text{Eve}} \{\text{Adv}^\text{p}(\text{Eve})\}$
  - Where we take the max over all Eve running in $t$ operations, making $q$ queries of $L$ bits to $E_k(.)$

**Security against Plaintext Recovery**
- We say $SE$ is $(t, q, l, \epsilon)$-secure against plaintext recovery if
  $$\text{Insec}^\text{p}(SE, t, q, l) \leq \epsilon$$
- Asymptotically, $SE$ is secure against plaintext recovery (PR-CPA) if for every polynomial time Eve, $\text{Adv}^\text{p}(\text{Eve})$ is negligible as a function of $k$.

---

**Problem with plaintext recovery**
- If Eve can reliably recover $m/2$ bits of the plaintext, she might be satisfied, and $SE$ would still be secure against plaintext recovery.
- Need a stronger definition, which is equivalent to the information-theoretic notion of not being able to learn a single bit about the plaintext.

**Indistinguishability under chosen plaintext attack**
- Define the oracle $LR_k(b, ..)$ as follows:
  $$LR(b, m_0, m_1) =$$
  - If $|m_0| \neq |m_1|$, return "-
  - Else return $E_k(m_b)$
- Suppose Eve is allowed to choose $m_0, m_1$. Then given $LR_k(b, ..)$ for randomly chosen $b$, she has one bit of uncertainty about $D_k(LR_k(b, m_0, m_1))$. 

**Indistinguishability under chosen plaintext attack**

In a *chosen plaintext attack*, Eve plays this game:

\[ \text{Exp}^{\text{cpa}}(b, \text{Eve}) = \]

Choose \( K \leftarrow U_k \)

Return \( \text{Eve}^{\text{LRK}}(b,\cdot,\cdot)(1^k) \).

Define the advantage of Eve, \( \text{Adv}^{\text{cpa}}(\text{Eve}) \), by

\[
\text{Pr}[\text{Exp}^{\text{cpa}}(1, \text{Eve}) = 1] - \text{Pr}[\text{Exp}^{\text{cpa}}(0, \text{Eve}) = 1]
\]

And \( \text{Insec}^{\text{cpa}}(\mathcal{SE}, t, q, l) = \max_{\text{Eve}}(\text{Adv}^{\text{cpa}}(\text{Eve})) \)

**IND-CPA is stronger than PR-CPA**

- Suppose we are given an Eve such that \( \text{Adv}^{\text{pr}}(\text{Eve}) \) is non-negligible. Then we will construct an IND-CPA adversary \( A \) which has

\[
\text{Adv}^{\text{cpa}}(A) \geq \text{Adv}^{\text{pr}}(\text{Eve}) - 1/2^m
\]

- This means that if we prove that \( \mathcal{SE} \) is IND-CPA then it is also PR-CPA.

**Example where PR-CPA is much weaker than IND-CPA**

- Suppose \( \mathcal{P} \) is a strong pseudorandom permutation family on \( \{0,1\}^* \). Let the message space be \( \{0,1\}^n \).

- Define the scheme \( \mathcal{E} = (G, E, D) \) as follows:

  - \( G(1^n) = \text{choose } K \leftarrow U_k \)
  - \( E_K(M) = F_K(M) \)
  - \( D_K(C) = F_K^{-1}(C) \).

- Claim: \( \text{Insec}^{\text{pr}}(\mathcal{ECB}, t, q, l) \leq \text{Insec}^{\text{pr}}(\mathcal{P}, t, q) + q2^{-\lambda} \)

- Yet \( \text{Insec}^{\text{cpa}}(\mathcal{ECB}, O(k), 2, 2k) = 1 \)
**IND-CPA encryption: CTR**

- Let $F_K : \{0,1\}^l \rightarrow \{0,1\}^l$ be a collection of pseudorandom functions.
- Define the stateful encryption scheme $CTR$ as follows:
  - $G(1^k) = \text{Choose } K \leftarrow U_k$
  - $E_K(m_0, m_1, \ldots, m_l) =$
    - Let $c_i = F_K(i) \oplus m_i$
    - update $j = j + l$
    - return $c_0, c_1, \ldots, c_l$
  - $D_K(c_0, c_1, \ldots, c_l) = E_K(c_0, c_1, \ldots, c_l)$

**IND-CPA security of $CTR$**

- Claim: Given any Eve which makes at most $q < 2^l$ queries of at most $\mu < l2^l$ bits, we can design a PRF Adversary $A$ with $Advprf(A) = \frac{1}{2} Advcpa(Eve)$.
- This gives us $Insecprf(CTR, t, q, \mu) \leq 2Insecprf(F, t, \mu/l)$
  - So if $F$ is a secure PRF then $CTR$ is IND-CPA

**Proof of claim**

- Given Eve, we define the PRF adversary $A$ as follows:
  - $A(1^k) =$
    - Choose $b \leftarrow U_1$.
  - Run Eve, responding to query $m_0, m_1, \ldots, m_l$ with $g(j) \oplus m_0, g(j+1) \oplus m_1, \ldots, g(j+l) \oplus m_l$ and updating $j$ appropriately.
  - If Eve outputs $b$, output 1, else output 0.

**Proof of CTR security**

- What is $Advprf(A)$?
  - First, notice that $Pr[A^{F_0(1^l)} = 1] = \frac{1}{2}$
    - If $g$ is a random function, then there is no correlation between the bit $b$ and the responses to Eve’s queries
  - Claim: $Pr[A^{F(1^l)} = 1] = \frac{1}{2} + \frac{1}{2} Advcpa(Eve)$
    - $Pr[A^b = 1 | b = 0] = Pr[Eve^{000\ldots} = 0]$ and $Pr[A^b = 1 | b = 1] = Pr[Eve^{111\ldots} = 1]$.
    - So $Pr[A^b = 1] = \frac{1}{2}(Pr[Eve^{000\ldots} = 0] + Pr[Eve^{111\ldots} = 1])$
    - $= \frac{1}{2}(1-Pr[Eve^{111\ldots} = 1]) + Pr[Eve^{111\ldots} = 1]$.
    - $= \frac{1}{2} + \frac{1}{2} Advcpa(Eve)$.

**Randomized (stateless) CTR**

- Define the scheme $RCTR$ as follows:
  - $G(1^k) = \text{Choose } K \leftarrow U_k$.
  - $E_K(r, m_0, m_1, \ldots, m_l) =$
    - Choose $r \leftarrow U_l$
    - Set $c_i = F_K(r+i) \oplus m_i$
    - Return $r, c_0, c_1, \ldots, c_l$
  - $D_K(r, c_0, c_1, \ldots, c_l) =$
    - Set $m_i = F_K(r+i) \oplus c_i$
    - Return $m_0, m_1, \ldots, m_l$.

**$RCTR$ is IND-CPA**

- Theorem: $Insecprf(RCTR, t, q, \mu) \leq 2Insecprf(F, t, \mu/l) + \mu q/l2^l$.
- Proof: Given an adversary Eve, define the PRF adversary $A$ as before. It still holds that when $A$ is given a pseudorandom oracle, it outputs 1 with probability $\frac{1}{2} + \frac{1}{2} Advprf(Eve)$. 
\[ \mathcal{CTR} \text{ is IND-CPA} \]

- It remains to bound the probability that \( A \) outputs 1 given a random function
  - If no input to the random function is repeated, then \( \Pr[A \text{ outputs } 1] = \frac{1}{2} \), as in previous argument.
  - If some input is repeated, \( A \) outputs 1 with probability at most 1. Call this event (a repeated input to the random function) COL.
  - So \( \Pr[A=1] \leq \frac{1}{2} + \Pr[\text{COL}] \)

\[ \text{Claim: } \Pr[\text{COL}] < q(\mu/l)2^{-l}. \]

- Notice that there are at most \((\mu/l)\) inputs to the random function.
- Let \( n_i \) be the number of inputs to \( f \) as a result of query \( l \).
- Suppose up to query \( i-1 \) there have been no repeated inputs to \( f \).
- What is the probability of a collision on query \( i \)?
- We get a collision with the \( j \)th query if \( r_j - n_i < r_i < r_j + n_j + 1 \), i.e., with probability \( n_i + n_j / 2 \).
- Thus the probability of collision on the \( i \)th query is at most \((i-1)n_i + n_i + n_2 + \ldots + n_{i-1})/2^l \).
- So the probability of a collision on any query is at most \( q(\mu/l)2^l \), as claimed.

\[ \Pr[\text{COL}] \]