Lecture Notes: 02/05/2002
Scribe: Elitza Maneva

Lecture Notes: 02/05/2002

In the previous lectures we presented several notions of security. We defined Semantic Security (SS), and Message Indistinguishability (IND) under Chosen Plaintext Attacks (CPA), and under Chosen Ciphertext Attacks (CCA1 and CCA2) . In this lecture we introduce the notion of message non-malleability (NM).

1 Motivation

Consider the following scenarios:

1. **Auction** To participate in an auction, the potential buyer sends an encoding of his name and the bid. This informnation is revealed at the end of the auction, either by sending it as a plaintext, or by publishing the secret key that was used to encode it. Refer to fig. 1. Suppose the buyer encodes his message with the Goldwasser-Micali cryptosystem. Then the two parts in the message are essentially coded separately (each bit for itself). An attacker can intercept the message, and even without knowing what the name was, he can replace it with an encoding of his own name, thus stealing the bid.

2. **Coin Flipping** Suppose Alice and Bob want to flip a coin. A possible protocol is shown on fig. 2. To change the result of the coin flip, an attacker would only need to know how to change the ciphertext, so that one of the bits in message r or s is flipped.

3. **Wired Equivalent Privacy (WEP)** A client sends the encoding of a TCP packet to the WEP base station, which decodes it, and sends the plaintext message to the server that is specified in the header of the packet (fig. 3). Upon receiving a valid packet, the server acknowledges with an ACK message. WEP uses the RC4 encryption algorithm, which is known as a stream cipher. A stream cipher operates by expanding a short key into an infinite pseudo-random key stream. The sender XORs the key stream with the plaintext to produce ciphertext. The receiver has a copy of the same key, and uses it to generate identical key stream. XORing the key stream with the ciphertext yields the original plaintext.

   An attacker can flip any two bits in the coded message by flipping those bits in the cipher text. The way the WEP integrity check protocol works, the server will acknowledge if and only if the bits in the position that were flipped are the same in the plain text. Thus the attacker can learn if any two bits are the same.

These examples motivate the definition of a new security requirement - message non-malleability (NM), which ensures that an adversary can not compute the encoding of a related message, given a ciphertext.

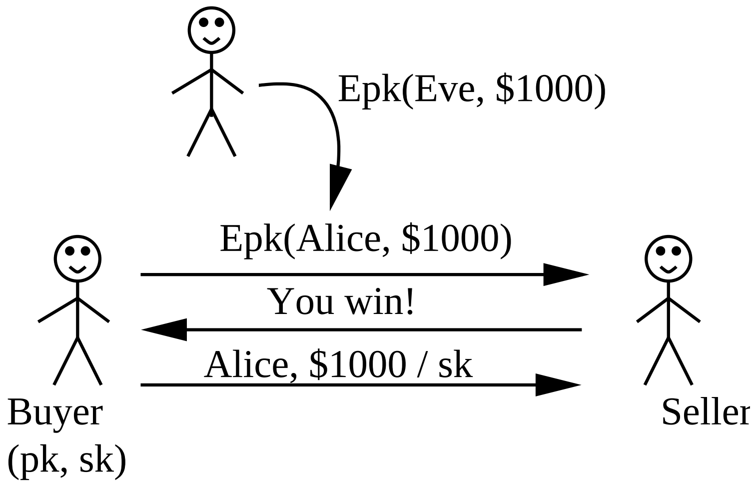

Figure 1: Protocol for holding an auction.

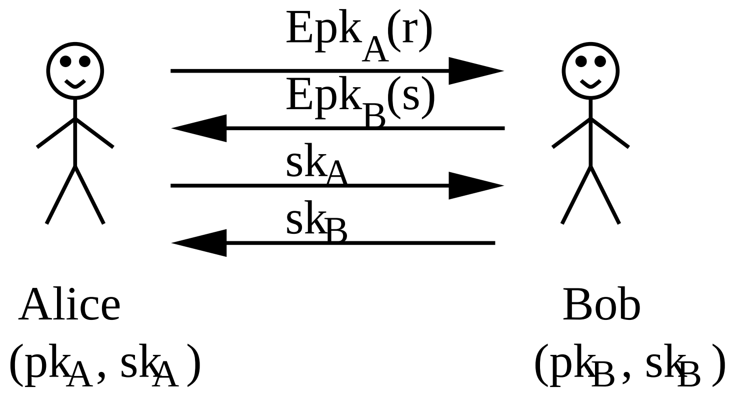

Figure 2: Protocol for coin flipping. If $r \oplus s = 0$ Alice wins, othersise Bob wins. This protocol is not secure in the message non-malleability sense.

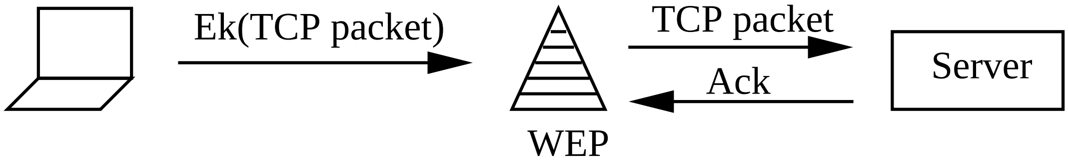

Figure 3: WEP protocol.

2 Definition of Security Under Message Non-Malleability

First recall the definition of IND-CPA and IND-CCA. We'll define those simultaneously.

DEFINITION 1 *A public-key scheme (E, D, G) is (T, q, ϵ)-secure in the IND-X sense if for all pairs of different messages m_0 and m_1 of the same length, and for every adversary A that runs in time t, and makes at most q quiries to oracle O*

$$\Pr_{(pk,sk)\in G}[A^O(pk, E_{pk}(m_0)) = 1] - \Pr_{(pk,sk)\in G}[A^O(pk, E_{pk}(m_1)) = 1] \leq \epsilon$$

where the oracle is

$$O = \begin{cases} D_{sk} & \text{if IND-CCA2} \\ - & \text{if IND-CPA} \end{cases}$$

and the adversary can't query the oracle at $E_{pk}(m_i)$.

Recall that the difference between CCA1 (a.k.a. "lunchtime' attack), and CCA2 is that the adversary in CCA1 does not have access to the decription oracle after he gets the ciphertext, but only before. The relationship between the security definitions we have so far is:

$$\text{Semantic Security} \Longleftrightarrow \text{IND-CPA} \longleftarrow \text{IND-CCA1} \longleftarrow \text{IND-CCA2}$$

DEFINITION 2 *A public-key scheme (E, D, G) is (T, q, ϵ)-secure in the NM-X sense if for all message distributions $\mathcal{M}$, and all relations $\mathcal{R} : \mathcal{M} \times \mathcal{M} \to \{0, 1\}$, and for every adversary A that runs in time t, and makes at most q quiries to oracle O, there exists another adversary A' that runs in time poly(t), such that*

$$\Pr_{(pk,sk)\in G, m\in\mathcal{M}}[\mathcal{R}(m, D_{sk}(A^O(pk, E_{pk}(m))] - \Pr_{(pk,sk)\in G, m\in\mathcal{M}}[\mathcal{R}(m, D_{sk}(A'(pk))] \leq \epsilon$$

where the oracle is

$$O = \begin{cases} D_{sk} & \text{if NM-CCA2} \\ - & \text{if NM-CPA} \end{cases}$$

and the adversary can't query the oracle at $E_{pk}(m)$.

This definition does not capture a situation in which the attacker is given more than one ciphertext and can compute a function of the original messages. This would be a problem, so we modify the relationship definition to $\mathcal{R} : \mathcal{M} \times \mathcal{M}^* \leftarrow \{0, 1\}$. The output of the adversary can be a vector of ciphertext, which does not include $E_{pk}(m)$.

3 Relation Between Security Under NM and Security Under IND

In this lecture we'll show the following relations:

$$\begin{array}{ccccc} & & \text{NM-CPA} & \leftarrow & \text{NM-CCA2} \\ & & \downarrow & & \downarrow \\ \text{Semantic Security} & = & \text{IND-CPA} & \leftarrow & \text{IND-CCA2} \end{array}$$

The horizontal implications are trivial, since the adversary would simply ignore the oracle. Next, we'll prove the vertical implications.

CLAIM 1
If a public-key scheme is (t, q, ϵ)-secure in NM-X sense, then it is $(t, q, 2\epsilon)$-secure in IND-X sense.

PROOF: Let's assume that the scheme is not secure in IND-X. Then it suffices to construct an adversary that breaks message non-malleability. There exist messages $m_0 \neq m_1$ and an adversary A^O, s.t.

$$\Pr_{(pk,sk)\in G}[A^O(pk, E_{pk}(m_0)) = 1] - \Pr_{(pk,sk)\in G}[A^O(pk, E_{pk}(m_1)) = 1] > 2\epsilon$$

We will use adversary $B^O(pk, c) = E_{pk}(m_{A^O(pk,c)})$ to show that the scheme is not secure under NM. Let the message distribution be the following:

$$\mathcal{M} = \begin{cases} m_0 & \text{with probability } 1/2, \\ m_1 & \text{with probability } 1/2. \end{cases}$$

and the relationship:

$$\mathcal{R}(u, v) = \begin{cases} \text{True} & \text{if } u = v, \\ \text{False} & \text{if } u \neq v. \end{cases}$$

We need to show that there exists no good simulator B', i.e. for all B' that run in time poly(t)

$$\Pr_{m\in\mathcal{M}}[\mathcal{R}(m, D_{sk}(B^O(pk, E_{pk}(m))))] - \Pr_{m\in\mathcal{M}}[\mathcal{R}(m, D_{sk}(B'(pk)))] > \epsilon. \tag{1}$$

Observe that the second term is

$$\Pr_{i\in\{0,1\}}[m_i = D_{sk}(B'(pk))] = \frac{1}{2}$$

The first term is

$$
\begin{aligned}
&\Pr_{i\in\{0,1\}}[m_i = D_{sk}(B^O(pk, E_{pk}(m_i))] \\
&= \Pr_{i\in\{0,1\}}[A^O(pk, E_{pk}(m_i)) = i] \\
&= \frac{1}{2}\Pr[A^O(pk, E_{pk}(m_0)) = 0] + \frac{1}{2}\Pr[A^O(pk, E_{pk}(m_1)) = 1] \\
&= \frac{1}{2}\Pr[A^O(pk, E_{pk}(m_0)) = 0] + \frac{1}{2}(1 - \Pr[A^O(pk, E_{pk}(m_1)) = 0] \\
&= \frac{1}{2} - \frac{1}{2}\mathbf{Adv}[\mathbf{A^O}]
\end{aligned}
$$

When we put these together, eq. 1 reduces to

$$
\Pr_{m\in\mathcal{M}}[\mathcal{R}(m, D_{sk}(B^O(pk, E_{pk}(m))))] - \Pr_{m\in\mathcal{M}}[\mathcal{R}(m, D_{sk}(B'(pk)))] = \frac{1}{2}\mathbf{Adv}[\mathbf{A^O}] > \epsilon
$$

This completes the proof. □