COM S 6830 – Cryptography October 27, 2011

Lecture 15: Public Key Encryption and Zero Knowledge

Instructor: Rafael Pass *Scribe: Sujay Jayakar (dsj36)*

1 Public Key Encryption

1.1 Preliminaries

Recall that a public key encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is secure if the following distributions are indistinguishable for all messages m_0 and m_1.

$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n), \mathsf{Enc}_{pk}(m_0)\}$$
$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n), \mathsf{Enc}_{pk}(m_1)\}$$

A trapdoor permutation is a family of functions $\{f_i(x)\}$ such that each f_i is a permutation and it is easy to sample a function from the family, sample each function's domain, and evaluate each function. Furthermore, each function is hard to invert in general, but easy to invert given some trapdoor information t. We saw previously that RSA is a trapdoor permutation with factorization as the trapdoor.

1.2 One-Bit Public Key Encryption

We will design a secure one-bit public key encryption scheme as follows. Let GenTDP return a randomly selected trapdoor permutation and $h(r)$ be a hardcore bit for that permutation.

$$\mathsf{Gen}(1^n) = \{(pk, sk) = (i, t) \leftarrow \mathsf{GenTDP}(1^n),\ \mathsf{return}\ (pk, sk)\}$$
$$\mathsf{Enc}_{pk}(m) = \{r \leftarrow \{0,1\}^n,\ \mathsf{return}\ f_{pk}(r) \mid\mid h(r) \oplus m\}$$
$$\mathsf{Dec}_{sk}(y \mid\mid c) = h(f_{pk}^{-1}(y)) \oplus c$$

Claim 1 *This one-bit public key encryption scheme is secure.*

Proof. Assume for contradiction there exists a nuPPT distinguisher $\mathcal{D}$ and a polynomial p such that for infinitely many n, $\mathcal{D}$ distinguishes the following distributions with probability greater than or equal to $1/p(n)$.

$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n : pk, \mathsf{Enc}_{pk}(0)\}$$
$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n : pk, \mathsf{Enc}_{pk}(1)\}$$

Note that since the only possible messages are zero or one, we do not have to quantify over all possible messages. We may rewrite these distributions, expanding out the definition

of Enc as follows.

$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n : pk, f_{pk}(r) \mid\mid h(r)\}$$
$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n : pk, f_{pk}(r) \mid\mid h(r) \oplus 1\}$$

By the prediction lemma, there exists a nuPPT attacker $\mathcal{A}$ that can tell from which distribution a sample originated. More formally, for infinitely many n,

$$\Pr[m \leftarrow \{0,1\}, (pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n : \mathcal{A}(pk, f_{pk}(r), h(r) \oplus m) = m] \geq \frac{1}{2} + \frac{1}{2p(n)}$$

Using $\mathcal{A}$, we will construct a new nuPPT algorithm $\mathcal{B}$ that contradicts the hardcore bit property of h. Define $\mathcal{B}$ as follows.

$$\mathcal{B}(pk, y) = \{c \leftarrow \{0,1\}, m = \mathcal{A}(pk, y, c), \ \mathsf{return}\ m \oplus c\}$$

Now, given the output $f_{pk}(r)$, we will show that $\mathcal{B}$ returns $h(r)$ with high probability.

$$\begin{aligned}
&\Pr[(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n, \mathcal{B}(pk, f_{pk}(r)) = h(r)] \\
&= \Pr[(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n, c \leftarrow \{0,1\}, \mathcal{A}(pk, f_{pk}(r), c) \oplus c = h(r)] \\
&= \Pr[(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n, m \leftarrow \{0,1\}, \mathcal{A}(pk, f_{pk}(r), m \oplus h(r)) \oplus m \oplus h(r) = h(r)] \\
&= \Pr[(pk, sk) \leftarrow \mathsf{Gen}(1^n), r \leftarrow \{0,1\}^n, m \leftarrow \{0,1\}, \mathcal{A}(pk, f_{pk}(r), m \oplus h(r)) = m] \geq \frac{1}{2} + \frac{1}{2p(n)}
\end{aligned}$$

The second step is just expanding the definition of $\mathcal{B}$, and the third uses the fact that if m is chosen randomly, $c \approx m \oplus h(r)$, by security of the one-time pad. Therefore, this argument contradicts our assumption that $h(r)$ was a hardcore bit of f_{pk}, so the scheme is secure. ■

Note that if we want to send multiple bits, we have to use a lot of randomness per bit. Each bit requires an n bit long random string, so we will need on the order of n^2 random bits to encrypt a message. This is entirely unnecessary. Use a similar scheme as the previous one, but alter the encryption function as follows.

$$\mathsf{Enc}_{pk}(m) = \{r \leftarrow \{0,1\}^n : \ \mathsf{return}\ f_{pk}^n(r) \mid\mid h(r) \oplus m_0 \mid\mid h(f_{pk}(r)) \oplus m_1 \mid\mid \ldots\}$$

In practice, users often use public key encryption to share a secret key for private key encryption, creating a "hybrid" encryption scheme.

2 Zero Knowledge

What does it mean for an interaction to not leak knowledge to another party? Consider the following toy examples. You are a journalist who has just learned of a murder, and everyone in the city knows of this murder. You would like to find out more about the

case, so you call the police department. However, the police department has a curious policy. If there has been a murder, they simply say, "There has been a murder," and hang up on you. As a second example, consider an even stranger police department. If there has been a murder, they flip a coin when the phone is ringing. If it is heads, they pick up, say, "There has been a murder," and hang up. If it is tails, they pick up and rudely hang up on you immediately.

In both cases the exchange with the police intuitively does not tell you anything: You already know there has been a murder, and that is precisely what the police station tells you! The key insight here is that you could simulate both of these conversations yourself without even calling the police. We will use this as our definition of zero knowledge. We will assume that polynomial time computation and flipping coins are cheap. In other words, our model of efficient computation will be PPT.

Definition 1 *An encryption scheme* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *is* **zero knowledge** *(ZK) if there exists a PPT algorithm* $\mathcal{S}$ *(for simulator) such that for all nuPPT distinguishers* $\mathcal{D}$*, there exists a negligible function* $\varepsilon(\cdot)$ *such that for all* $n \in \mathbb{N}$ *and* $m \in \{0,1\}^n$*,* $\mathcal{D}$ *distinguishes the following distributions with probability less than or equal to* $\varepsilon(n)$*.*

$$\{k \leftarrow \mathsf{Gen}(1^n) : \mathsf{Enc}_k(m)\}$$
$$\{S(1^n)\}$$

In the case of public key encryption, we may analogously define the distributions as

$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n) : pk, \mathsf{Enc}_{pk}(m)\}$$
$$\{(pk, sk) \leftarrow \mathsf{Gen}(1^n) : pk, S(1^n, pk)\}.$$

Sometimes an equivalent "behavior" definition is better suited. Here, we represent an attacker's previous knowledge as a polynomially long bit string z.

Definition 2 *An encryption scheme is also zero knowledge secure if for all nuPPT attackers* $\mathcal{A}$*, there exists a PPT simulator* $\mathcal{S}$ *such that for all nuPPT distinguishers* $\mathcal{D}$*, there exists a negligible function* $\varepsilon(\cdot)$ *and a polynomial* p *such that for all* $n \in \mathbb{N}$*,* $m \in \{0,1\}^n$*, and* $z \in \{0,1\}^{p(n)}$*,* $\mathcal{D}$ *distinguishes the following distributions with probability less than or equal to* $\varepsilon(n)$*.*

$$\{k \leftarrow \mathsf{Gen}(1^n), \mathcal{A}(z, \mathsf{Enc}_k(m))\}$$
$$\{S(1^n, z)\}$$

Although zero knowledge security intuitively seems to be much stronger than our traditional definition of security, they are, indeed, equivalent. A sketch of a proof that zero knowledge implies secure would be to notice that $\mathsf{Enc}(m_0) \approx S(1^n) \approx \mathsf{Enc}(m_1)$, so the hybrid lemma tells us $\mathsf{Enc}(m_0) \approx \mathsf{Enc}(m_1)$. Assuming traditional security, we could just set $\mathcal{S}$ to pick a random key and output the encryption of some arbitrary string.