6.875/18.875 Cryptography and Cryptanalysis February 23, 2005

Handout 6: Equivalence of GM and Semantic Security

The TA's came up with a simple proof that if a cryptosystem is GM-secure, it is also semantically secure. Below is their original write-up of this proof.

Notation: Note that in this write-up, the cryptosystem in question is denoted as $\mathcal{C}$, the key-generation algorithm is also denoted as $\mathcal{C}$ and the public key is denoted by E. Furthermore, when E is a public key, the notation $E(m)$ is used to denote the encryption of message m using public key E. (This notation is quite natural if you think of the key-generation procedure as producing the code of the encryption algorithm with the public-key hard-coded in.)

GM Security $\implies$ Semantic Security

We show that $\neg$ Semantic Security $\implies$ $\neg$ GM-Security. Let $\{M_n\}$ be message spaces, f be a polynomial-time computable function, and $\{A_n\}$ be circuits such that for a fixed $c > 0$ and infinitely many n

$$\Pr[A_n(E,\alpha) = f(m) \mid m \leftarrow M_n, E \leftarrow \mathcal{C}(1^n), \alpha \leftarrow E(m)] \geq \tilde{p} + \frac{1}{n^c} \tag{1}$$

where $\tilde{p} = \mathbb{E}_{E \leftarrow \mathcal{C}(1^n)}[p_E]$ is the expected prediction probability without the knowledge of α.

Consider the following algorithm $T_n : (E, m_0, m_1, \alpha) \rightarrow \{0,1\}$.

1. Let $\beta \leftarrow A_n(E,\alpha)$.
2. If $\beta = f(m_0)$ but $\beta \neq f(m_1)$, output 0.
3. If $\beta = f(m_1)$ but $\beta \neq f(m_0)$, output 1.
4. Otherwise, output a random value from $\{0,1\}$ with probability $\frac{1}{2}$ each.

The test is very intuitive. We simply run A_n on the challenge α. Since we expect A_n to correctly predict the value of f, we compare its output β with $f(m_0)$ and $f(m_1)$. Note that the test is clearly polynomial time since all the steps (including computations of f) are polynomial time.

If exactly one of the tests succeed, we output the corresponding message. Otherwise, we flip a coin as we did not learn anything. For specific m_0 and m_1, let

$$q(m_0, m_1) = \Pr[T_n(E, m_0, m_1, \alpha) = i \mid i \in_r \{0,1\}, E \leftarrow \mathcal{C}(1^n), \alpha \leftarrow E(m_i)]$$

be the probability that T_n distinguishes encryptions of m_0 and m_1.

We show that T_n violates the GM-security of $\mathcal{C}$ by finding two particular messages m_0 and m_1 that are distinguished by T_n, i.e. $q(m_0, m_1) \geq \frac{1}{2} + \frac{1}{2n^c}$ (same c as in (1)). To show the existence of such m_0 and m_1 we use the probabilistic method. We pick both m_0 and m_1 *independently at random according to the given probability distribution* M_n (that violates the Semantic Security in (1)). We then argue that T_n has non-negligible *expected* advantage in distinguishing a random encryption of m_0 or m_1, i.e. $q := \mathbb{E}_{m_0,m_1}[q(m_0, m_1)] \geq \frac{1}{2} + \frac{1}{2n^c}$. Hence, the required m_0 and m_1 exist.

It remains to prove the bound on q. We note that since the algorithm T_n is symmetric in m_0 and m_1, q equals to the expected probability that T_n outputs 0 if α is an encryption of m_0, i.e. without loss of generality we can assume that $i = 0$. Now, our experiment can be viewed as the following. Pick $m_0 \leftarrow M_n$, $E \leftarrow \mathcal{C}(1^n)$, $\alpha \leftarrow E(m_0)$, $\beta \leftarrow A_n(E, \alpha)$. Now we pick a *brand new* message $m_1 \leftarrow M_n$ and run steps 2–4 of T_n. q is the probability that we output 0. Before computing q, we claim that

$$\Pr[\beta = f(m_0)] \geq \tilde{p} + \frac{1}{n^c}; \qquad \Pr[\beta = f(m_1)] \leq \tilde{p} \tag{2}$$

Indeed, the first bound follows directly from (1), as $\beta \leftarrow A_n(E, \alpha)$ and $\alpha \leftarrow E(m_0)$. For the second bound, we observe that for any *fixed* E, the message m_1 is chosen *independent* of m_0, $\alpha \leftarrow E(m_0)$ and, therefore, $\beta \leftarrow A_n(E, \alpha)$. Hence, for any fixed E the probability that $f(m_1)$ equals to β is at most the probability that it equals to any pre-specified element, which is at most p_E. Since for a fixed E, our probability is stochastically dominated by p_E, we can take the expectation over E to obtain the claimed bound.

Now, using the fact $\Pr[A \wedge B] + \Pr[A \wedge \bar{B}] = \Pr[A]$, we can compute the probability q of outputting 0 in the following way:

$$\begin{aligned}
q &= \Pr[\beta = f(m_0) \wedge \beta \neq f(m_1)] + \frac{1}{2}(\Pr[\beta = f(m_0) = f(m_1)] + \Pr[\beta \notin \{f(m_0), f(m_1)\}]) \\
&= \frac{1}{2}\left(\Pr[\beta = f(m_0) \wedge \beta \neq f(m_1)] + \Pr[\beta = f(m_0) \wedge \beta = f(m_1)]\right) + \\
&\quad \frac{1}{2}\left(\Pr[\beta = f(m_0) \wedge \beta \neq f(m_1)] + \Pr[\beta \neq f(m_0) \wedge \beta \neq f(m_1)]\right) \\
&= \frac{1}{2}\left(\Pr[\beta = f(m_0)] + \Pr[\beta \neq f(m_1)]\right) = \frac{1}{2} + \frac{1}{2}\left(\Pr[\beta = f(m_0)] - \Pr[\beta = f(m_1)]\right) \\
&\overset{(2)}{\geq} \frac{1}{2} + \frac{1}{2}\left((\tilde{p} + \frac{1}{n^c}) - \tilde{p}\right) = \frac{1}{2} + \frac{1}{2n^c}
\end{aligned}$$

This concludes the proof. $\square$