CS 6222 Grad Crypto September 16, 2025

Lecture 7: Chosen Plaintext Attacks and CPA-Secure Encryption

Lecturer: Jack Doerner *Scribe: Raul Hernandez*

1 Topics Covered

- The eavesdropping game and EAV-security.
- Chosen plaintext attacks (CPA) and IND-CPA-security.
- IND-CPA-secure encryption.

2 The eavesdropping game and EAV-security

Today, we are finally going to build encryption. In a previous lecture, we defined single-message EAV1-security. This definition quantified over all messages and all NUPPT adversaries, which could potentially have the single message hardcoded. Now we will extend this definition to consider multiple messages, and we will explicitly give the adversary the ability not just to know but to choose the messages.

Definition 1 (The Eavesdropping Game)**.** *The game* $\mathsf{EAV}_b^{\Pi,\mathcal{A}}$ *for any two-part adversary* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ *and any symmetric encryption scheme* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *is as follows:*

1. $k \leftarrow \mathsf{Gen}(1^n)$ *generate a key.*
2. $\left(\vec{m}^0, \vec{m}^1, s\right) \leftarrow \mathcal{A}_1(1^n)$ *such that* $\left|\vec{m}^0\right| = \left|\vec{m}^1\right|$ *and for all* $i \in [t]$*, we have that* $|m_i^0| = |m_i^1|$ *where* $\vec{m}^b = \left(m_1^b, \cdots, m_t^b\right)$*.*
3. *For all* $i \in [t]$*, encrypt* $c_i \leftarrow \mathsf{Enc}_k(m_i^b)$*.*
4. *Output* $\mathcal{A}_2\left(s, \vec{c}\right)$*.*

Definition 2A (EAV-security). *A symmetric-key encryption (*SKE*) scheme* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *has indistinguishable encryptions under eavesdropping (*EAV*-security) if for all* NUPPT $\mathcal{A}$*, there exists negligible* ε *such that for all* $n \in \mathbb{N}$

$$\left|\Pr\left[\mathsf{EAV}_b^{\Pi,\mathcal{A}}(n) = b : b \leftarrow \{0,1\}\right] - \frac{1}{2}\right| < \varepsilon(n)$$

Definition 2B (EAV-security (equivalent)). *As above, but for all* NUPPT $\mathcal{A}$

$$\left\{\mathsf{EAV}_0^{\Pi,\mathcal{A}}(n)\right\}_{n\in\mathbb{N}} \approx_c \left\{\mathsf{EAV}_1^{\Pi,\mathcal{A}}(n)\right\}_{n\in\mathbb{N}}$$

3 Chosen plaintext attacks (CPA) and IND-CPA-security

Now consider the scenario that we have some message m, and we compute $c_1 \leftarrow \mathsf{Enc}_k(m)$ and $c_2 \leftarrow \mathsf{Enc}_k(c_1)$. Notice that the above game tells us nothing about this scenario: it is not possible for any message in the EAVgame to depend upon a ciphertext, except by chance. Nevertheless, this is a realistic scenario [1] We can modify our game to capture scenarios like this one by allowing the adversary to choose plaintexts *adaptively* based upon all previous ciphertexts. This gives us the Chosen Plaintext Attack game:

Definition 3 (The CPA Indistinguishability Game)**.** *The game* $\mathsf{IND\text{-}CPA}_b^{\Pi,\mathcal{A}}(n)$ *for any two-part adversary* NUPPT $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ *and any symmetric encryption scheme* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *is as follows:*

1. $k \leftarrow \mathsf{Gen}(1^n)$.
2. $(m_0, m_1, s) \leftarrow \mathcal{A}_1^{\mathsf{Enc}_k(\cdot)}(1^n)$.
3. $c^* \leftarrow \mathsf{Enc}_k(m_b; r^*) : r^* \leftarrow$ *randomness domain of* Enc_k.[2]
4. *Output* $\mathcal{A}_2^{\mathsf{Enc}_k(\cdot)}(s, c^*)$.

Definition 4A (IND-CPA-security). *A symmetric-key encryption (*SKE*) scheme* $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *has indistinguishable ciphertexts under chosen plaintext attacks (*IND-CPA*-security) if for all* NUPPT $\mathcal{A}$*, there exists negligible* ε *such that for all* $n \in \mathbb{N}$

$$\left| \Pr\left[\mathsf{IND\text{-}CPA}_b^{\Pi,\mathcal{A}}(n) = b : b \leftarrow \{0,1\} \right] - \frac{1}{2} \right| < \varepsilon(n)$$

Definition 4B (IND-CPA-security (equivalent)). *As above, but for all* NUPPT $\mathcal{A}$

$$\left\{ \mathsf{IND\text{-}CPA}_0^{\Pi,\mathcal{A}}(n) \right\}_{n\in\mathbb{N}} \approx_c \left\{ \mathsf{IND\text{-}CPA}_1^{\Pi,\mathcal{A}}(n) \right\}_{n\in\mathbb{N}}$$

Almost all practical encryption schemes in use today are (at least putatively) IND-CPA-secure, and this is usually the minimal notion of security on which we insist for encryption. Note, however, that stronger security notions exist.

4 Constructing CPA-secure encryption

Note 1. *For convenience, let* $F_{k,1^\ell} : \{0,1\}^{|k|} \to \{0,1\}^\ell$ *is a* PRF*with parametric output length. Such an object can be constructed by combining any length-preserving PRF with a PRG.*

Construction 1 (Encryption from a PRF for messages of any polynomially-bounded length)**.**

- $\mathsf{Gen} : 1^n \mapsto k : k \leftarrow \{0,1\}^n$.

[1] It would be quite unfortunate if encrypting a message twice made it *less* secure!

[2] Here we give a name to the random coins used to encrypt m_b, so that we can refer to them later.

- $\mathsf{Enc} : k, m \mapsto r \| \left(m \oplus F_{k,1^{|m|}}(r)\right) : r \leftarrow \{0,1\}^n$
- $\mathsf{Dec} : k, c \mapsto \left(c' \oplus F_{k,1^{|m|}}(r)\right) : r \| c' := c$

Theorem 1. *If $F_{k,1^\ell}$ is a* PRF *then Construction 1 is* IND-CPA*-secure.*

Proof. Consider an inefficient variant of Construction 1 based on random functions:

Construction 2 ($\widetilde{\Pi} = \left(\widetilde{\mathsf{Gen}}, \widetilde{\mathsf{Enc}}, \widetilde{\mathsf{Dec}}\right)$)**.**

- $\widetilde{\mathsf{Gen}} : 1^n \mapsto \{f_{1^\ell} : f_{1^\ell} \leftarrow \mathcal{F}_{n,\ell}\}_{\ell \in \mathbb{N}}$ [3]
- $\widetilde{\mathsf{Enc}} : \{f_{1^\ell}\}_{\ell \in \mathbb{N}}, m \mapsto r || m \oplus f_{1^{|m|}}(r) : r \leftarrow \{0,1\}^n$.
- $\widetilde{\mathsf{Dec}} : \{f_{1^\ell}\}_{\ell \in \mathbb{N}}, c \mapsto c' \oplus f_{1^{|c'|}}(r) : c := r || c'$.

If we let

$$\mathcal{H}^{\mathcal{A}}_{0,b} = \left\{\mathsf{IND\text{-}CPA}^{\Pi,\mathcal{A}}_b(n)\right\}_{n\in\mathbb{N}} \qquad \text{and} \qquad \mathcal{H}^{\mathcal{A}}_{1,b} = \left\{\mathsf{IND\text{-}CPA}^{\widetilde{\Pi},\mathcal{A}}_b(n)\right\}_{n\in\mathbb{N}}$$

then using Lemma 1 and Lemma 2 below, as well as the hybrid lemma (which implies transitivity for computational indistinguishability), we can show that $\mathcal{H}^{\mathcal{A}}_{0,0} \approx_c \mathcal{H}^{\mathcal{A}}_{0,1}$. □

Lemma 1. *The* PRF*-security of $F_{k,1^\ell}$ implies that for all* NUPPT *$\mathcal{A}$ and $b \in \{0,1\}$, we have $\mathcal{H}^{\mathcal{A}}_{0,b} \approx_c \mathcal{H}^{\mathcal{A}}_{1,b}$.*

Proof. Consider the reduction $R_b^{O(\cdot,\cdot)}(1^n)$ with access to an oracle $O : 1^\ell \times \{0,1\}^n \to \{0,1\}^\ell$

Construction 3 ($R_b^{O(\cdot,\cdot)}(1^n)$)**.**

1. *Let* $\widehat{\mathsf{Enc}}_n : m \mapsto r \| m \oplus O\left(1^{|m|}, r\right) : r \leftarrow \{0,1\}^n$.
2. $R_b^{O(\cdot,\cdot)}$ *emulates* $(m_0, m_1, s) \leftarrow \mathcal{A}_1^{\widehat{\mathsf{Enc}}_n(\cdot)}(1^n)$ *internally.*
3. $R_b^{O(\cdot,\cdot)}$ *computes* $c \leftarrow \widehat{\mathsf{Enc}}_n(m_b)$.
4. $R_b^{O(\cdot,\cdot)}$ *outputs* $\mathcal{A}_2^{\widehat{\mathsf{Enc}}_n(\cdot)}(c, s)$.

Claim 1. $R_b^{F_{k,(\cdot)}(\cdot)} : k \leftarrow \{0,1\}^n \equiv \mathsf{IND\text{-}CPA}^{\Pi,\mathcal{A}}_b(1^n)$

This is true because $O(\cdot,\cdot) = F_{k,(\cdot)}(\cdot)$ implies

$$\begin{aligned}\widehat{\mathsf{Enc}}_n(m) &\equiv r \| m \oplus F_{k,1^{|m|}}(r) : r \leftarrow \{0,1\}^n, k \leftarrow \{0,1\}^n \\ &\equiv \mathsf{Enc}_k(m) : k \leftarrow \mathsf{Gen}(1^n)\end{aligned}$$

[3] Notice that keys are *infinitely* long, and even the descriptions of the individual random functions that make up a key are exponentially long, relative to the security parameter. We will only use this scheme for a thought experiment, so this will not be a problem for us.

Claim 2. $R_b^{f_{(\cdot)}(\cdot)}(1^n) : \{f_{1^\ell} : f_{1^\ell} \leftarrow \mathcal{F}_{n,\ell}\}_{\ell\in\mathbb{N}} \equiv \mathsf{IND\text{-}CPA}_b^{\widetilde{\Pi},\ell}(1^n)$.

This second claim is analogous to the previous claim, and it can be shown in a similar way. Finally, by the PRF-security of F and the closure of computational indistinguishability under NUPPTpostprocessing, we have

$$\left\{R_b^{f_{(\cdot)}(\cdot)}(1^n) : \{f_{1^\ell} : f_{1^\ell} \leftarrow \mathcal{F}_{n,\ell}\}_{\ell\in\mathbb{N}}\right\}_{n\in\mathbb{N}} \approx_c \left\{R_b^{F_{k,(\cdot)}(\cdot)}(1^n) : k \leftarrow \{0,1\}^n\right\}_{n\in\mathbb{N}} \qquad \square$$

Lemma 2. *For all* NUPPT $\mathcal{A}$*, we have that* $\mathcal{H}_{1,0}^{\mathcal{A}} \approx_c \mathcal{H}_{1,1}^{\mathcal{A}}$.

Proof. Recall that in $\mathcal{H}_{1,b}^{\mathcal{A}}$, $\mathcal{A}$ has oracle access to $\widetilde{\mathsf{Enc}}_{\{f\}}(\cdot)$. Let S be the set of r values used by the oracle $\widetilde{\mathsf{Enc}}$ in responding to queries in either of these hybrids.

Claim 3. *Since* $\mathcal{A}$ *is* NUPPT*, there is a polynomial* p *such that* $|S| < p(n)$ *in the context of* $\mathcal{H}_{1,b}^{\mathcal{A}}$ *for* $b \in \{0,1\}$.

Claim 4. *There exists some negligible* ε *such that in the context of* $\mathcal{H}_{1,b}^{\mathcal{A}}$ *for* $b \in \{0,1\}$*, we have* $\Pr[r^* \in S] = \frac{|S|}{2^n} < \varepsilon(n)$.[4]

Claim 5. *For all* NUPPT $\mathcal{A}$ *and* $n \in \mathbb{N}$*, we have that*

$$\Pr\left[\mathsf{IND\text{-}CPA}_0^{\widetilde{\Pi},\mathcal{A}}(n) = 1 \mid r^* \notin S\right] \equiv \Pr\left[\mathsf{IND\text{-}CPA}_1^{\widetilde{\Pi},\mathcal{A}}(n) = 1 \mid r^* \notin S\right]$$

The above claim holds because in both the left and right-hand experiments, c^* is completely uniform from the point of view of the adversary. Recall that $\mathcal{H}_{1,b}(n) = \mathsf{IND\text{-}CPA}_b^{\widetilde{\Pi},\mathcal{A}}(n)$. For all $n \in \mathbb{N}$, we have that

$$\begin{aligned}
\Pr[\mathcal{H}_{1,b}(n) = 1] &= \Pr[\mathcal{H}_{1,b}(n) = 1 \wedge r^* \in S] + \Pr[\mathcal{H}_{1,b}(n) = 1 \wedge r^* \notin S] \\
&\leq \Pr[r^* \in S] + \Pr[\mathcal{H}_{1,b}(n) = 1 \mid r^* \notin S] \cdot \Pr[r^* \notin S] \\
&= \Pr[r^* \in S] + \Pr[\mathcal{H}_{1,1-b}(n) = 1 \mid r^* \notin S] \cdot \Pr[r^* \notin S] \quad \text{(by Claim 5)} \\
&= \Pr[r^* \in S] + \Pr[\mathcal{H}_{1,1-b}(n) = 1 \wedge r^* \notin S] \\
&\leq \Pr[r^* \in S] + \Pr[\mathcal{H}_{1,1-b}(n) = 1] \\
&< \Pr[\mathcal{H}_{1,1-b}(n) = 1] + \varepsilon(n) \quad \text{(by Claim 4)}
\end{aligned}$$

Thus, for all NUPPT $\mathcal{A}$, there exists a negligible ε such that for all $n \in \mathbb{N}$, we have that

$$\left|\Pr\left[\mathsf{IND\text{-}CPA}_0^{\widetilde{\Pi},\mathcal{A}}(n) = 1\right] - \Pr\left[\mathsf{IND\text{-}CPA}_1^{\widetilde{\Pi},\mathcal{A}}(n) = 1\right]\right| < \varepsilon(n) \qquad \square$$

[4]Recall that r^* is the randomness used by the $\mathsf{IND\text{-}CPA}$-game itself to encrypt the challenge message m_b.