CS 120/CSCI E-177: Introduction to Cryptography

Salil Vadhan and Alon Rosen Sept. 21 & 26, 2006

Lecture Notes 3:

Private-Key Encryption: Perfect Secrecy

Recommended Reading.

- Katz–Lindell, Chapter 2.

1 Private-Key (aka Symmetric) Encryption

- The setting for private-key encryption is the following: two parties share a *secret key* and want to exchange messages *privately* over "insecure channel". For now, we will not worry about how they came to share the secret key.
- Kerchkhoff's Principle: Assume encryption/decryption algorithms are known to adversary. Only thing secret is the *key*.
- For now, "insecure channel" means that adversary can *listen* to all messages sent, but cannot inject/alter messages, i.e. *passive* rather than *active*.
- **Definition 1** *A* (private-key) encryption scheme *consists of three algorithms* (G, E, D)*, as follows:*
  - *The* key generation *algorithm* G *is a randomized algorithm that returns a* key $k \in \mathcal{K}$*; we write* $k \xleftarrow{R} G$*.*
  - *The* encryption algorithm E *is a randomized algorithm that takes a key* $k \in \mathcal{K}$ *and a* plaintext *(aka* message*)* $m \in \mathcal{P}$ *and outputs a ciphertext* $c \in \mathcal{C}$*; we write* $c \xleftarrow{R} E_k(m)$*.*
  - *The* decryption algorithm D *is a deterministic algorithm that takes a key* $k \in \mathcal{K}$ *and a ciphertext* $c \in \mathcal{C}$ *and returns a plaintext* $m \in \mathcal{P}$*.*

  The *message space* $\mathcal{P}$ is often the set of strings of a given length. The ciphertext space $\mathcal{C}$ does not have to equal the plaintext space. We require $D_k(E_k(m)) = m$ for all $m \in \mathcal{P}$.
- The definition describes the functionalities of the encryption scheme but does not take security into account yet. For example:
- Examples:
  - Shift cipher (cf. Caesar cipher). The key is a random number: $k \xleftarrow{R} \{0, \ldots, 25\}$, the message space is $\mathcal{P} = \{A, \ldots, Z\}^\ell$ (strings of length ℓ over the English alphabet) so we can see the message as $m \in \{0, \ldots, 25\}^\ell$. $E_k(m_1 m_2 \cdots m_\ell) = c_1 c_2 \cdots c_\ell$, where $c_i = m_i + k \pmod{26}$.

- Substitution cipher. The key k is a random permutation of $\{0, \ldots, 25\}$. $E_k(m_1 m_2 \cdots m_\ell) = k(m_1)k(m_2)\cdots k(m_\ell)$.
- One-time pad. The message space consists of binary strings of length ℓ and the key k is a random element of $\{0,1\}^\ell$. $E_k(m) = m \oplus k$ (bitwise XOR). The decryption is $D_k(c) = c \oplus k$.

2 Perfect Secrecy

- What does it mean for something to be secret? How to define security? Some attempts:
  - Adversary can't determine key from ciphertext.
  - Adversary can't determine plaintext.
  - Adversary can't determine any symbol of plaintext.
  - Adversary can't determine "any information" about plaintext.
- **Definition 2 (perfect indistinguishability)** *Encryption scheme satisfies* perfect indistinguishability *if for every $m_1, m_2 \in \mathcal{P}$ and $K \xleftarrow{R} G$, the random variables $E_K(m_1)$ and $E_K(m_2)$ have the same distribution. That is, for every c,*

$$\Pr[E_K(m_1) = c] = \Pr[E_K(m_2) = c],$$

where the probabilities are taken over $k \xleftarrow{R} G$ and the coin tosses of E.

Idea: the adversary sees the same distribution of ciphertext, regardless of the message sent. Note that there is no probability distribution over the messages; rather we assume that the adversary knows the possible messages in advance. Intuitively, the case of two messages is the worst case (the adversary knows all but "one bit" of information in advance), and hence is representative of the security of an encryption scheme.

- **Proposition 3** *Shift and Substitution ciphers do not satisfy perfect indistinguishability for messages of length > 1.*

Proof:

- **Proposition 4** *One-time pad satisfies perfect indistinguishability.*

Proof:

- **Definition 5 (Shannon secrecy)** *Let M be a distribution on $\mathcal{P}$. An encryption scheme satisfies* Shannon secrecy *with respect to M if for every $m \in \mathcal{P}$ and every $c \in \mathcal{C}$,*

$$\Pr[M = m | E_K(M) = c] = \Pr[M = m]$$

where the probabilities are taken over $K \xleftarrow{R} G$, M, *and the coin tosses of* E.

Idea: after seeing the ciphertext, the adversary doesn't know more about the message than before seeing the ciphertext; the a posteriori knowledge is the same as the a priori knowledge. We think of M as known to the adversary.

- **Proposition 6** *An encryption scheme satisfies perfect indistinguishability if and only if it satisfies Shannon secrecy (with respect to any* M *s.t.* $\Pr[M = m] > 0$ *for all* $m \in \mathcal{P}$*). Thus we refer to both as* perfect secrecy *(or* perfect security*).*

  Proof: We only prove that perfect indistinguishability implies Shannon secrecy. The converse is Lemma 2.2 in Katz–Lindell. By Bayes' Law,

$$\Pr[M = m | E_K(M) = c] = \frac{\Pr[E_K(M) = c | M = m] \cdot \Pr[M = m]}{\Pr[E_K(M) = c]}$$

  We need to prove that $\Pr[E_K(M) = c | M = m] = \Pr[E_K(M) = c]$, i.e. $\Pr[E_K(m) = c] = \Pr[E_K(M) = c]$. This follows from perfect indistinguishability. ■

- Why isn't this course over?

- **Theorem 7** *If an encryption scheme is perfectly secure, then the number of keys is at least the size of the plaintext space.*

  Proof:

- How to get around this limitation? We can relax our definition of secrecy:
- "Statistical" security: only require encryptions of all messages to be statistically close.
  - Let X and Y be random variables taking values in a set S. X and Y are called *statistically* ε*-indistinguishable* if for every event $T \subseteq S$

$$|\Pr[X \in T] - \Pr[Y \in T]| \leq \varepsilon.$$

    T is also called a *statistical test.*
  - **Definition 8 (statistical secrecy)** *Encryption scheme satisfies* statistical ε-indistinguishability *if for every two* $m_1, m_2 \in \mathcal{P}$*, the random variables* $E_K(m_1)$ *and* $E_K(m_2)$ *are statistically* ε*-indistinguishable. (These random variables are taken over* $K \xleftarrow{R} G$ *and the coin tosses of* E*.)*
  - Intuitively, adversary has probability at most ε of getting information about the plaintext.
  - Insufficient to go beyond the barrier we have with Shannon secrecy: requires $|\mathcal{K}| \geq (1 - \varepsilon) \cdot |\mathcal{P}|$.

- "Computational" security: only protect against adversaries with *limited computational resources*, i.e. efficient adversaries with a reasonable amount of computational power $\Rightarrow$ REST OF THIS COURSE.
- Other communication settings — quantum cryptography, beacon of random bits,...