COM S 6830 – Cryptography Aug 30, 2011

Lecture 2: Perfect Secrecy

Instructor: Rafael Pass *Scribe: Sidharth Telang*

1 Overview

In this lecture we explore what it means for a symmetric encryption scheme to be secure. We define perfect secrecy introduced by Shannon in his work on information theory and observe that the Caesar and substitution cipher introduced in the last lecture are not secure under this notion. We modify these ciphers so as to use different keys for every letter of the message and extend this idea to bit strings to define the One-time pad. We prove this encyption scheme to be perfectly secret.

However, we notice that the One-time pad requires keys that are of the same length as the message. We proceed to prove that any encryption scheme that is perfectly secret cannot have a key-space smaller than its message-space, which implies the One-time pad is the best one can hope for in terms of minimizing the lengths of keys while maintaining perfect secrecy. The workaround to this problem is to assume adversaries are computationally bounded, *i.e.* they are modeled by what we define as probabilistic polynomial time algorithms.

2 Perfect secrecy, One-time pad

We review the definition of a symmetric encryption scheme

Definition 1 *A symmetric encryption scheme is a tuple of algorithms* (Gen, Enc, Dec) *with message-space* M *and key-space* K *where* Gen *and* Enc *are possibly randomized and* Dec *is deterministic such that for all messages* $m \in M$ *and keys* $k \in K$

$$Pr[Dec_k(Enc_k(m)) = m] = 1$$

Such a scheme is used by say Alice to send Bob a message $m \in M$ in the presence of an eavesdropper Eve in the following way

- Alice and Bob share a key k generated by Gen

$$k \leftarrow Gen$$

- Alice computes ciphertext c on m and k using Enc

$$c \leftarrow Enc_k(m)$$

- Bob decrypts c under k using Dec

$$m' = Dec_k(c)$$

By the above definition we have $m' = m$.

The most basic notion of security would require that the attacker Eve not learn the key, which along with Kirchoff's principle would imply that Gen be a randomized algorithm. To enhance this notion, we could require that the attacker not be able to recover any partial information about the message from the ciphertext. This requirement is captured in the following definition of perfect secrecy.

Definition 2 *An encryption scheme (Gen, Enc, Dec) over message-space M and key-space K is said to be perfectly secret if for all messages $m_1, m_2 \in M$ and all ciphertexts c we have*

$$Pr[k \leftarrow Gen : Enc_k(m_1) = c] = Pr[k \leftarrow Gen : Enc_k(m_2) = c]$$

In other words, the attacker cannot even get any partial information about the message from the ciphertext since all messages give identical distibutions on the ciphertext.

We can see that the Substitution cipher is not perfectly secret, as a ciphertext with repeated letters (obtainable from a message with repeated letters) cannot be obtained from a message with distinct letters. However, it is perfectly secure, if a randomly generated key is used for each letter of the message. We extend this idea to bit strings as follows.

Let the message-space M and key-space K be $\{0,1\}^n$.

- Gen: $k \leftarrow K$
- $Enc_k(m)$: $c = m \oplus k$ where $\oplus$ denotes bitwise XOR
- $Dec_k(c)$: $m = c \oplus k$

Definition 3 *We define the One-time pad as the above symmetric encryption scheme.*

Proposition 1 *The One-time pad is perfectly secure.*

Proof. Consider any messages $m_1, m_2 \in M$ and ciphertext c. If $c \in \{0,1\}^n$ then $Pr[k \leftarrow Gen : Enc_k(m_1) = c] = Pr[k \leftarrow Gen : k = m_1 \oplus c] = 2^{-n} = Pr[k \leftarrow Gen : Enc_k(m_2) = c]$. If $c \notin \{0,1\}^n$ then $Pr[k \leftarrow Gen : Enc_k(m_1) = c] = Pr[k \leftarrow Gen : Enc_k(m_2) = c] = 0$.

Note that the One-time pad uses keys of the same length as the messages. Unfortunately, this is an implication of perfect secrecy by the following theorem by Shannon.

Theorem 1 *For all perfectly secret encryption algorithms* (Gen, Enc, Dec) *on message-space* M *and key-space* K, $|K| \geq |M|$.

Proof. Assuming $|K| < |M|$ we shall derive a contradiction to perfect secrecy. Consider a message $m_0 \in M$ and key $k_0 \in K$ in the range of Gen. Let $c \leftarrow Enc_{k_0}(m_0)$. Consider the set $S = \{m \in M : \exists k \in K \; s.t. \; Dec_k(c) = m\}$. Since Dec is deterministic, $|S| \leq |K| < |M|$. Therefore there exists $m_1 \in M$ such that $Pr[k \leftarrow Gen : Enc_k(m_1) = c] = 0$, else c is obtainable from m_1 but doesn't decrypt to it, violating the definition of symmetric encryption schemes. Since $Pr[k \leftarrow Gen : Enc_k(m_0) = c] > 0$ we have a contradiction.

Hence to use keys that are shorter than messages, we need to relax the notion of perfect secrecy. One way of doing this while maintaining some sense of security is assuming that attackers are computationally bounded *i.e.* they are capable of only efficient computation and the key-space is big enough to rule out brute force attacks. We model efficient computation by probabilistic polynomial time (PPT) algorithms.

Definition 4 *A probabilistic polynomial time (PPT) algorithm is an algorithm with access to an infintely long random tape, which for all inputs* $x \in \{0,1\}^*$ *and random tapes halts within* $p(|x|)$ *steps for some polynomial* p. *A PPT algorithm* A *is said to compute* $f : \{0,1\}^* \to \{0,1\}^*$ *with probability* p *if for all inputs* $x \in \{0,1\}^*$ *we have*

$$Pr[A(x) = f(x)] \geq p$$