CSIT 495/595 - Introduction to Cryptography
Perfect Secrecy

Bharath K. Samanthula
Department of Computer Science
Montclair State University

Security Definition 1

- Computational Security
  - Assuming that Malory has limited computational resources, it will be infeasible for Malory to infer anything useful from the communication between Alice and Bob
  - In practice, we will prove that if a certain problem is hard (e.g. factoring large integers) than breaking a certain cryptographic primitive will be computationally infeasible (also known as provable security)

Security Definition 2

- Unconditional Security (i.e. Perfect Security)
  - Even if Malory has infinite amount of computational resources, he cannot learn anything from the communication
- Pros: Better Protection compared Computational Security
- Cons: Secret keys have to be as large as the message size

Probability - Overview

A discrete random variable **X** is defined by specifying

- A finite set X
  (e.g. the possible values a tossed dice can take.)
- A probability distribution on X such that
  the probability of **X** takes on the value x
  is denoted as $Pr[\mathbf{X} = x]$ (e.g. the probability that
  we get tails after a coin flip)

If **X** is fixed define $Pr[\mathbf{X} = x]$ as $Pr[x]$

$Pr[x] >= 0$ for all $x \in X$

$(\sum_{x \in X} Pr[x]) = 1$

Probability - Overview

Given an event $E \subset X$, define
$Pr[x \in E] = \sum_{x \in E} Pr[x]$

Example:

- Random variable **Z**: result of throwing a pair of dice
- Defined on set $Z = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$
- Define event S_4 as the sum of the dices is 4.
- $S_4 = \{(1,3),(2,2),(3,1)\}$
- $Pr[S_4] = 1/12$

Probability - Overview

- Given two random variables **X** and **Y**
  - $Pr[x, y]$ is the joint probability
  - $Pr[x|y]$ is the conditional probability
- Random variables **X** and **Y** are independent if
  - $Pr[x, y] = Pr[x].Pr[y]$
- $Pr[x, y] = Pr[x|y].Pr[y]$
- Bayes Theorem
  - If $Pr[y] > 0$ then $Pr[x|y] = \frac{Pr[y|x].Pr[x]}{Pr[y]}$

Shift Cipher - Probability Analysis

Example:

- Let K and M denote the random variables denoting the key and message used such that $\Pr[K = k] = 1/26$, $\Pr[M = a] = 0.7$ and $\Pr[M = z] = 0.3$
- What is the probability that the ciphertext is B?
- Two possible cases: ($M = a$ and $K = 1$) or ($M = z$ and $K = 2$)
- $\Pr[M = a \wedge \ K = 1] = 0.7 * (1/26)$ and $\Pr[M = z \wedge \ K = 2] = 0.3 * (1/26)$ (Note: K and M are independent)
- $\Pr[C = B] = \Pr[M = a \wedge \ K = 1]$ + $\Pr[M = z \wedge \ K = 2]$ = 1/26
- $\Pr[M = a | C = B] = \frac{Pr[c=B|M=a]*Pr[M=a]}{Pr[C=B]} = 0.7$

Perfect Secrecy - Formal Definition

- An encryption scheme (i,e., Gen, Enc, Dec) is perfectly secure if

$$\Pr[M = m | C = c] = \Pr[M = m], \forall\ m \in \mathcal{M} \text{ and } c \in \mathcal{C}$$

- This definition states that a posteriori probability that the plaintext is m given that ciphertext is c is equal to the a priori probability that the plaintext is m
- Probability distribution of the ciphertext does not depend on the plaintext
- Formally, for any two messages $m, m' \in \mathcal{M}$ and $c \in \mathcal{C}$:

$$\Pr[\mathrm{Enc}_k(m) = c] = \Pr[\mathrm{Enc}_k(m') = c]$$

One-Time Pad

- Vernam patented this scheme in 1917
- Fixes the vulnerabilities of Vigenere Cipher by using very long keys
- $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$
- Encryption: bitwise exclusive OR of m and k, $c \leftarrow m \oplus k$
- Decryption: $m \leftarrow c \oplus k$
- Example: Let $m = 00101$ and $k = 10010$. What is c?

One-Time Pad: Perfect Secrecy Proof

- One-Time pad encryption scheme is Perfectly Secret
- Need to show that $\Pr[m|c] = \Pr[m]$ for one-time pad
- Proof:

$$
\begin{aligned}
Pr[M = m|C = c] &= \frac{Pr[C = c|M = m] * Pr[M = m]}{Pr[C = c]} \\
&= \frac{Pr[K = m \oplus c] * Pr[M = m]}{Pr[C = c]} \\
&= \frac{2^{-\ell} * Pr[M = m]}{\sum_{m \in M} Pr[C = c|M = m] * Pr[M = m]} \\
&= \frac{2^{-\ell} * Pr[M = m]}{2^{-\ell} * \sum_{m \in M} Pr[M = m]} \\
&= Pr[M = m]
\end{aligned}
$$

One-Time Pad: Limitations

- Widely used in mid-20th century (e.g., red phone linking White House and the Kremlin during the cold war)
- Rarely used now-a-days due to many limitations
- Key is as long as the message
  - limits the use of the scheme in case of very large messages
  - Sometimes parties cannot predict an upper bound for the message size in advance
- Is secure if only used once (with the same key) – Why??
  - If same key is used, then $c \oplus c' = m \oplus m'$!!

Shannon's Theorem

- Let $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$. An encryption scheme is perfectly secure iff:
  - Every $k \in \mathcal{K}$ is chosen with equal probability by Gen
  - For every m, c, there exists a unique key such that $\mathsf{Enc}_k(m) = c$
- For perfect secrecy, we must have $|\mathcal{K}| \geq |\mathcal{C}| \geq |\mathcal{M}|$ (can you see why??)

Summary

- Computational Security vs. Unconditional Security
- Definition of Perfect Secrecy
- One-Time pad and its limitations
- Shannon's Theorem

Useful References

- Chapter 2, Introduction to Modern Cryptography by Jonathan Katz and Yehuda Lindell, 2nd Edition, CRC Press, 2015.
- http://www.ics.uci.edu/~stasio/fall04/lect1.pdf
- http://www.cs.umd.edu/~jkatz/crypto/f02/lectures/lecture3.pdf