Week 1 **CS6160: Cryptology**

History, Encryption Systems, and Perfect Security

1 History

1.1 The Casesar and Mono-alphabetic Substitution Ciphers

According to history, Julius Caesar (who lived around 50BC) used to encode any messages that he sent, so that anyone who intercepts the message cannot read it. The method he used was to shift each character by 3 letters further in the alphabet. For example, the sentence "I can read this." would be encoded as "L fdq uhdg wklv." Clearly, this scheme is easy to decrypt even if one doesn't know the shift, by simply tring the 26 possibilities.

Relatively stronger, but still very insecure, is the monoalphabetic substitution cipher, in which each character is represented by some other character or symbol, not necessarily by a common shift. For example, the ciphertext "Uv jqz vor frzu el vukrz; uv jqz vor jeyzv el vukrs." could be an encoding of the message "It was the best of times; it was the worst of times." where the substitution mapping is given by the table below.

Character	a	b	e	f	h	i	m	o	r	s	t	w
Substitution	q	f	r	l	o	u	k	e	y	z	v	j

Such ciphers were analyzed as early as 800 AD by the Arab mathematician Al Kindi who wrote a treatise on cryptanalysis. The method of solving such a cipher is to do a frequency analysis, i.e. to count the occurrences of each character. The frequency distribution in large texts in a natural language tend to be similar to that observed from statistical data. For the English language, the frequency distribution is as shown in the table below.

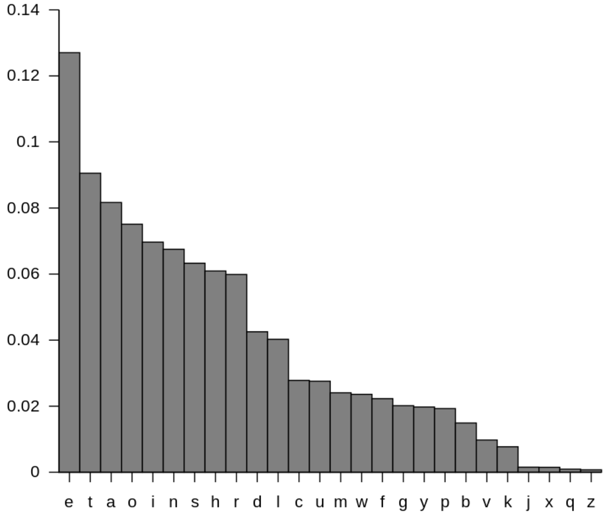

Figure 1: Frequency distribution of English letters, shown as probabilities. Courtesy Wikipedia

1.2 The Vigenere Cipher

Polyalphabetic ciphers attempt to mask the frequency distribution, and an example of this is the Vigenere Cipher, invented in the 15th century.

In the Vigenere Cipher, the secret key is a string usually English. To encode a message, the key is repeated till its length becomes equal to that of the message and then corresponding characters are added modulo 26. An example should make this clear. Let the message be m=ATTACKATDAWN and let the key be k=RING. Then Enc(m,k) is obtained by adding ATTACKATDAWN and RINGRINGRING to obtain: RBGGTSNZUIJT. Here, the letters A,B,…,Z are mapped to 0,1,…,25. Thus, T+I is (19+8) mod 27=1, which corresponds to B.

While the Vigenere Cipher is stronger than a simple substitution cipher, it can also be broken. Methods to break it were found independendently by Kasiski and Babbage in the 19th century, and are based on the following idea: When two patterns repeat, then the difference between their positions is a likely candidate for a multiple of the key length. For example, if we observe the ciphertext QZVUABAVILQZVU then the pattern QZVU appears in two positions that are 6 apart, we may guess that the key length is 2 or 3 or 6.

For a candidate key-length L, frequency analysis of characters in positions that are a multiple of L apart will help find the key.

1.3 Modern Cryptography

Modern cryptography has many applications: Internet and emails, banking, credit cards, file encryption, content protection (eg: DRMs) and many more. It uses a variety of tools to solve a variety of security problems. Some examples:

- Confidentiality: This is the most basic requirement, that a message accessed by an eavesdropper should not reveal information.
- Integrity: A message should not be tampered by an attacker.
- Authentication: The message comes from the source that it claims to come from. Also authentication of users in any system.

Something that characterizes classical cryptosystems is that they were designed in an ad-hoc manner and have no security guarantees. Most of them were initially thought to be strong, but broken later. How has modern cryptography learnt from these failures? Here are some typical features that characterize modern cryptography.

- Formal definitions of security. We cannot design a good cryptosystem without knowing what kind of security guarantees we are aiming for. We'll see examples from the next class and more examples later.
- Proofs of security. Many modern cryptographic constructions and protocols come with proofs of what kind of security guarantees they are able to offer.
- Precise assumptions. Proofs of security are not unconditional, i.e. they often use some assumptions about the underlying primitives used.
- Use of primitives. Apart from individual/isolated constructions, many protocols make use of a set of standard primitives, for example one-way functions, pseudo-random generators, pseudo-random permutations etc. These primitives are assumed to have certain security guarantees which enables the designer to combine them in ways that make apparent the security guarantee of the larger system.

2 Private-Key Encryption Systems

Definition 1. *Given a message-space M, cipher-text space M, and a key-space K,* **a private-key encryption system** Π *over* (M, C, K) *is a triple of algorithms* (Gen, Enc, Dec) *with the following properties:*

- *Gen is called the key-generation algorithm and outputs a random element of K.*
- *Enc is a function $Enc : M \times K \to C$ and is called the encryption algorithm.*
- *Dec is a function $Dec : C \times K \to M$ and is called the decryption algorithm.*
- *$Enc(k, m) = c$ implies $Dec(c, k) = m$.*

Imagine that a trusted third-party runs the key-generation algorithm Gen() and distributes a common key k to Alice and Bob. When Alice wants to send a message m to Bob, she computes $c = Enc(m, k)$ and sends c to Bob. Bob can now decrypt the message by finding $m = Dec(c, k)$.

3 Perfect Security

In the 1950s, Claude Shannon introduced the idea of perfect secrecy/perfect security. Intuitively, a cryptosystem is perfectly secure if **zero information** about the message is obtained from observing its ciphertext. This is made precise as follows: consider a probability distribution on M. Then after observing a given ciphertext, the posterior distribution on M should be the same as the prior distribution. For example, if the message could have been any message in M with uniform probability, then after observing the ciphertext c, it should be the case that c could be the encryption of any message in M with uniform probability.

Formally, we have the following definition.

Definition 2. *A private-key cryptosystem $\Pi = (Gen, Enc, Dec)$ is perfectly secure if for all $m \in M, c \in C$ and for all random variables X defined on M, we have the following:*

$$Pr[X = m | Enc(X, k) = c] = Pr[X = m].$$

The substitution cipher is not perfectly secure by the above definition; fix n and consider the substitution cipher to be defined on $M = \Sigma^3$, where Σ consists of the English alphabet (upper and lower case). The key k is a permutation of Σ generated at random. Let X be a random variable that is uniformly distributed on M, and let $c = ABA$. Then for $m = ABC$, we have: $Pr[X = m] = \frac{1}{52^3}$, but $Pr[X = m|Enc(X, k) = c] = 0$.

An example of a perfectly secure cryptosystem is the One Time Pad, described by Vernam in 1914. In the One Time Pad, M is of the form $M = \Sigma^n$ for some alphabet Σ, and $K = C = \Sigma^n$. Suppose that $|\Sigma| = L$. For computation, we identify Σ with $\{0, 1, \ldots, L - 1\}$; the encryption function is:

$$Enc(m, k) = (m + k) \pmod{\mathrm{L}}$$

where the addition is co-ordinate wise, i.e. for each character separately. The decryption function is:

$$Dec(c, k) = (c - k) \pmod{\mathrm{L}}$$

Two examples:

- Let $\Sigma = \{A, \ldots, Z\}^6$ and $k = ABCABZ$. Then $Enc(SECRET, k) = TGFSGS$.
- Let $\Sigma = \{0, 1\}^6$ and $k = 101100$. Then $Enc(011001, k) = 110101$.

Note that for the binary alphabet, the One-Time Pad Encryption function is the same as the XOR of the message and the key.

3.1 Proof that the One-Time Pad is perfectly secure

Let X be a random variable taking values in M. Let m be an arbitrary message. We have:

$$\begin{aligned} Pr[X = m|Enc(X, k) = c] &= \frac{Pr[X = m \wedge Enc(X, k) = c]}{Pr[Enc(X, k) = c]} \\ &= Pr[X = m]\frac{Pr[Enc(X, k) = c|X = m]}{Pr[Enc(X, k) = c]} \\ &= Pr[X = m] \end{aligned}$$

The first and second equality use Bayes' theorem that $Pr[A|B] = \dfrac{Pr[A \cap B]}{Pr[B]}$. The last equality comes from the fact that for every fixed message m, we have: $Pr[Enc(m,k) = c] = \dfrac{1}{L^n}$ since the key k is picked uniformly at random. This also shows that $Pr[Enc(X,k) = c] = \dfrac{1}{L^n}$ for any r.v. X defined on M so that both numerator and denominator have the same value.

4 Two More Definitions of Perfect Security

We now give two more definitions of perfect security; it turns out that all three definitions are equivalent. That is, a cryptosystem is secure with respect to one definition if and only if it is secure with respect to each of the other two.

Definition 3. *A private-key cryptosystem* $\Pi = (Gen, Enc, Dec)$ *is perfectly secure if for all* $m_0, m_1 \in M, c \in C$*, we have:*

$$Pr[Enc(m_0,k) = c] = Pr[Enc(m_1,k) = c].$$

Intuitively, the above definition is a converse of the first definition; it says that the distribution of the ciphertext is independent of the message.

Given a private-key cryptosystem $\Pi = (Gen, Enc, Dec)$ over (M, C, K) and an adversary A, the **indistinguishability experiment** $PrivK^{Eav}_{A,\Pi}$ is defined as follows.

- Alice uses $Gen()$ to pick a key $k \in K$ uniformly at random.
- The adversary picks two messages m_0, m_1 from M and sends them to Alice.
- Alice picks $b \in \{0,1\}$ uniformly at random. Then Alice computes $c = Enc(m_b, k)$ and sends c to the adversary.
- The adversary outputs a value $b' \in \{0,1\}$ which is their guess of b.
- The output of the experiment, also denoted by $PrivK^{Eav}_{A,\Pi}$, is defined to be 1 if $b' = b$ and 0 otherwise.

The goal in the above game is for the adversary to output $b' = b$ with as large a probability as possible. Note that by simply guessing randomly, the adversary will guess correctly with probability 1/2. Perfect indistinguishability is the property that no adversary can do better. Formally, we have the following definition.

Definition 4. *A cryptosystem* $\Pi = (Gen, Enc, Dec)$ *over* (M, C, K) *is* **perfectly indistinguishable** *if for every adversary* A*, it holds that*

$$Pr[PrivK^{Eav}_{A,\Pi} = 1] = \frac{1}{2}.$$

5 Class Exercises

1. Show that the One-Time Pad is perfectly secure according to Definition 3.

   Solution: For all m, c, we have: $Pr[Enc(m,k) = c] = \frac{1}{L^n}$.

2. Show that the following scheme is not perfectly secure according to Definition 4. We define $M = \{0,1\}^n$, $K = \{0,1\}^{n-1}$ and $C = \{0,1\}^n$. Given a message $m = m_1 m_2 \ldots m_n$ and $k = k_1 \ldots k_{n-1}$, we define $Enc(m,k) = ((m_1 \ldots m_{n-1}) XOR (k_1 \ldots k_{n-1})) || m_n$. Here, $||$ is the symbol for concatenation.

   Solution: The adversary chooses two messages such that one of them has last bit zero and the other message has last bit one. On seeing the ciphertext $c = c_1 \ldots c_n$, the adversary outputs the value of c_n and thus correctly finds the message corresponding to c with probability 1.

3. Suppose that an adversary is able to guess a secret key k correctly with probability $\frac{1}{1000}$. Show that if the adversary makes 3000 such guesses independently, then the probability that at least one guess is correct is greater than 0.9.

   Solution: The probability that all guesses are incorrect is $(1-\frac{1}{1000})^{3000} < e^{-3} < 0.1$, thus the probability that at least one guess is correct is greater than 0.9. Here, we used the inequality $1 - x < e^{-x}$ which is useful in such calculations.