COMP3260 Data Security

Lecture 3

Prof Ljiljana Brankovic

Lecture Overview

1. Rate of the Language
2. Redundancy
3. Equivocation
4. Perfect Secrecy
5. One-Time Pad
6. Symmetric Cipher Model
7. Kerckhoffs' Laws
8. Codes and Ciphers

"Cryptography and Data Security" by D. Denning [2]

Note that in-text references and quotes are omitted for clarity of the slides. When you write as essay or a report it is very important that you use both in-text references and quotes where appropriate.

Rate of the Language

Consider a language L consisting of messages of N characters.

The **rate** r **of the language** L is the average entropy per character, that is,

$$r = \frac{H(X)}{N}$$

Real languages consist of messages of varying lengths.

In that case we can define the rate r_N of the language for messages of length N.

Rate of the Language

As N increases, r_N tends to a constant r which is then the **rate of the language**.

For large N, estimates of r for English range from $1.0\ \frac{bit}{letter}$ to $1.5\,\frac{bit}{letter}$.

We shall use $1.5\,\frac{bit}{letter}$ as the estimate for English in our calculations.

Absolute Rate of the Language

If the alphabet of L consists of L characters then the absolute rate R of the language L is $R = \log_2 L$. The absolute rate of English is $R = \log_2 26 = 4.7 \frac{bit}{letter}$.

The absolute rate is the maximum entropy of the characters under any probability distribution. That is, the absolute rate is the maximum number of bits of information that could be encoded in each character assuming all possible sequences of characters are equally likely.

If all sequences of characters in a language have the same probability then $r = R$.

Redundancy

The absolute rate of English is significantly greater that the actual rate because English is highly redundant.

Mst ids cn b xpresd n fwr ltrs, bt th xprnc s mst nplsnt!

In any natural language, as well as in programming languages, redundancy arises from the structure of the language. The redundancy is reflected in the statistical properties of actual meaningful messages:

- single letter frequency distribution
- diagram frequency distribution
- trigram frequency distribution
- N-gram frequency distribution

Redundancy

As longer sequences are considered, the proportion of meaningful messages to the total number of possible letter sequences decreases.

In practice, the rate of language (entropy per character) is determined by estimating the entropy of N-grams for increasing values of N.

As N increases, the entropy per character decreases because there are fewer choices and some choices are much more likely than others.

The rate of language is estimated by extrapolating for large N.

Redundancy

The **redundancy** D of a language with rate r and absolute rate R is defined as

$$D = R - r$$

For English,

$$D = 4.7 - 1.5 = 3.2$$

Thus, English is 68% redundant, since $\frac{D}{R} = 0.68$.

When using the rate of $1 \frac{bit}{letter}$ it is around 79% redundant.

Equivocation

The uncertainty of a message can be further reduced given additional information.

Example 1: Suppose X is a 32-bit integer, all values equally likely so $H(X) = 32$. Suppose we learn that X is even. How much does this additional information reduce the entropy of X?

By 1 bit because all even integers have 0 as their last bit.

Equivocation

The entropy of a message X, given some additional information Y, is measured by the **equivocation** $H_Y(X)$,the uncertainty about X given knowledge of Y.

The **equivocation** $H_Y(X)$ is the **conditional entropy** of X given Y:

$$H_Y(X) = -\sum_{X,Y} p(X,Y)\log_2 p_Y(X) =$$

$$= \sum_{X,Y} p(X,Y)\log_2 \frac{1}{p_Y(X)} =$$

$$= \sum_{Y} p(Y) \sum_{X} p_Y(X)\log_2 \frac{1}{p_Y(X)}$$

Equivocation

$p_Y(X)$ is the conditional probability of message X given message Y and $p(X,Y)$ is the joint probability of message X and message Y:

$$p(X,Y) = p_Y(X)p(Y)$$

If events X and Y are independent, then $p_Y(X) = p(X)$ and we have

$$p(X,Y) = p(X)p(Y)$$

and

$$H_Y(X) = H(X)\sum_Y p(Y) = H(X)$$

Equivocation

Example 2: Let $n = 4$ and $p(X) = \frac{1}{4}$ for each message X so $H(X) = \log_2 4 = 2$. Let $m = 4$ and $p(Y) = \frac{1}{4}$ for each message Y. Suppose each message Y narrows down the choice of X to two of the four messages, both equally likely:

$$Y_1\text{: } X_1 \; or \; X_2$$
$$Y_2\text{: } X_2 \; or \; X_3$$
$$Y_3\text{: } X_3 \; or \; X_4$$
$$Y_4\text{: } X_4 \; or \; X_1$$

What is the equivocation of X given Y?

In this case the knowledge of Y reduces the uncertainty of X to 1 bit.

Perfect Secrecy

Shannon studied the information theoretic properties of cryptographic systems in terms of three classes of information:

- plaintext messages M occurring with known probabilities $p(M)$, where $\Sigma_M\, p(M) = 1$;
- ciphertext messages C occurring with known probabilities $p(C)$, where $\Sigma_C\, p(C) = 1$;
- keys K chosen with probabilities $p(K)$, where $\Sigma_K\, p(K) = 1$.

Perfect secrecy is defined by the condition

$$p_C(M) = p(M)$$

where $p_C(M)$ is the probability that M was sent given that C was received.

Perfect Secrecy

The probability of receiving C given that M was sent is the sum of the probabilities p(K) of the keys K that encipher M as C:

$$p_M(C) = \sum_{K, E_K(M)=C} p(K)$$

A necessary and sufficient condition for perfect secrecy is that for every C and for all M, $p_M(C)=p(C)$.

Perfect Secrecy

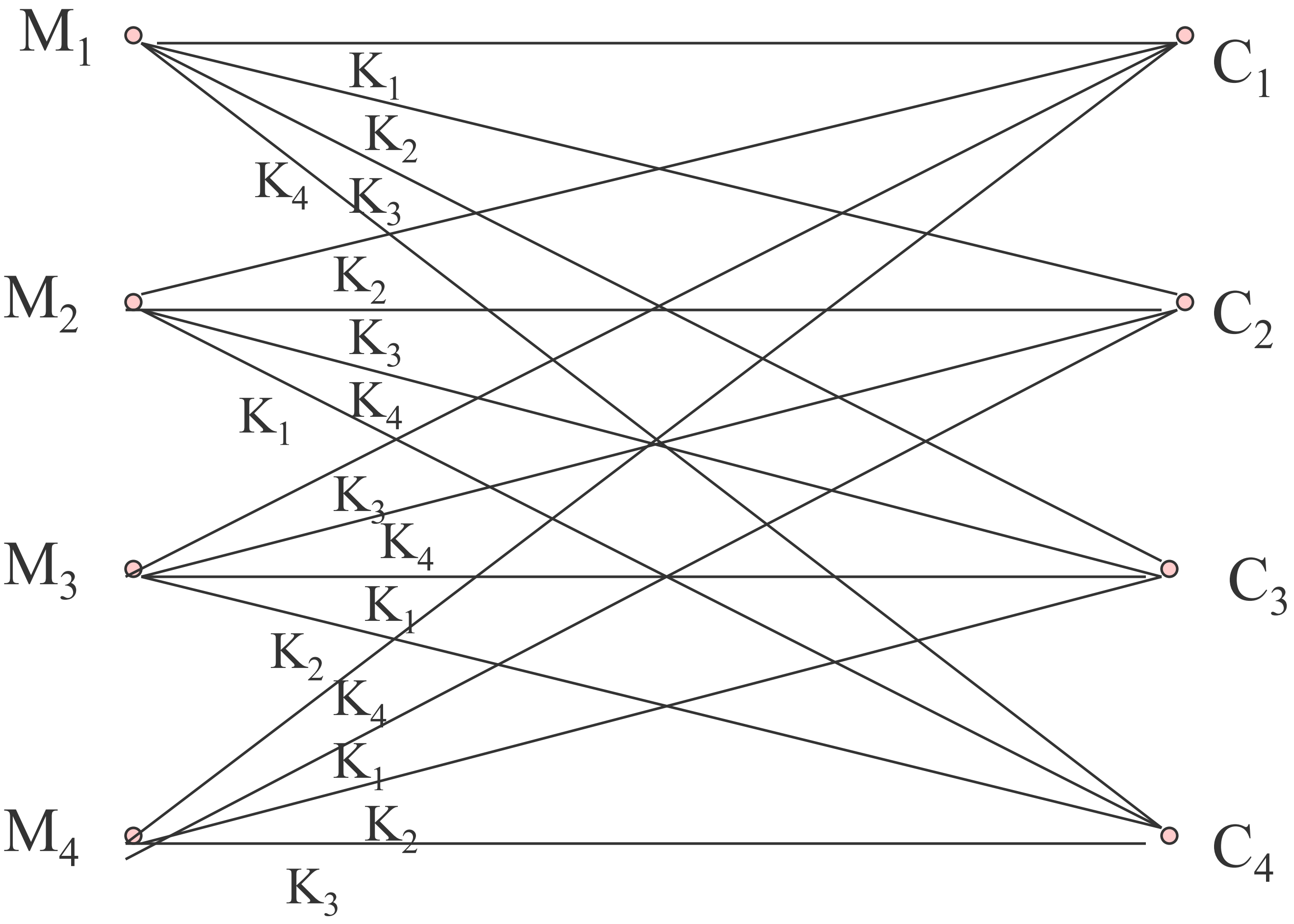
So the probability of receiving a particular ciphertext C given that M was sent is the same as the probability of receiving C given that some other (any other) message M was sent.

Perfect secrecy is possible using a completely random key at least as long as the message it enciphers.

Perfect secrecy requires that the number of keys must be at least as great as the number of possible messages.

Otherwise there would be some message M such that for a given C, no K deciphers C into M , implying that $p_C(M)=0$. However, M is not impossible, so $p_C(M) \neq p(M)$.

Perfect Secrecy



Perfect Secrecy

Example: Suppose we intercept a ciphertext which was produced by a *Caesar* cipher, with key K.

C=DOXDRYECKXNWOXGYEVNXYDLOOXYEQR

Is this cipher perfectly secure?

No. We cannot achieve perfect secrecy because the number of possible keys is smaller than the number of possible English sentences of length 31. This cipher is easily broken because only one of the possible 26 keys (K=10) produces a meaningful message: **TENTHOUSANDMENWOULDNOTBEENOUGH**

Perfect Secrecy

We have $p_C(M)$=1 and $p_C(M')$=0 for every other message M'.

$p_M(C)$=p(10)=1/26 and $p_M'(C)$=0 for every other message M'.

$p_C(M)$ is certainly greater than p(M) and $p_M(C)$ is greater that p(C).

One-Time Pad

Modification of the preceding example to achieve perfect secrecy: shift each letter not by a constant number of places but by a random number.

Then $K=k_1k_2\ldots$, where each k_i is a random integer in the range [0,25].

Perfect secrecy is achieved since any 31 character long message could be enciphered to C.

One-Time Pad

The ciphertext in the last example could have resulted from using the key K =3,3,4,22,3,4,24,21,22,10,9,10,14,10,20,16,24,14,10,14,11,18, 20,18,19,22,14,13
to encrypt the message

M'=**ALTHOUGHONEMANMIGHTJUSTSUFFICE**

A cipher that uses a non-repeating random key stream is called a **one-time pad**.

One time pads are the only ciphers that achieve perfect secrecy.

Unicity Distance

Shannon measured the secrecy of a cipher in terms of the key equivocation $H_C(K)$ for a given ciphertext C:

$$H_C(K) = \sum_C p(C) \sum_K p_C(K) \log_2 \frac{1}{p_C(K)}$$

where $p_C(K)$ is the probability of K given C.

If $H_C(K)=0$ then there is no uncertainty and the cipher is theoretically breakable given enough resources.

As the length of the ciphertext increases, the equivocation usually decreases.

Unicity Distance

The **unicity distance** is the smallest N such that $H_C(K)$ is close to 0, that is, it is the amount of ciphertext needed to uniquely determine the key.

A cipher is **unconditionally secure** if $H_C(K)$ never approaches 0, regardless how large N is.

Most ciphers are too complex to determine the probabilities needed to calculate the unicity distance.

Approximating Unicity Distance

However, for ciphers which are close to a random cipher model it is possible to find a good approximation of the unicity distance.

Assume that each plaintext and ciphertext message comes from a finite alphabet of L symbols.

Then there are 2^{RN} possible messages of length N, where $R=\log_2 L$.

These messages are partitioned into two subsets:

- a set of 2^{rN} **meaningful** messages, each with probability $(1/2^{rN})=2^{-rN}$
- a set of $2^{RN}-2^{rN}$ **meaningless** messages, each with probability 0

Approximating Unicity Distance

Assume that there are $2^{H(K)}$ keys, where H(K) is the key entropy, i.e., the number of bits in the key.

The key entropy is also called the **entropy of the system**: it is a measure of the size of the key space ***K***, approximated by $H(K)=\log_2|\boldsymbol{K}|$.

For example, a cryptosystem with a 64-bit key has an entropy of 64 bits.

The probability of each key is $p(K)=1/2^{H(K)}=2^{-H(K)}$

Approximating Unicity Distance

A **random cipher** is a cipher in which for each key K and ciphertext C, the decipherment $D_K(C)$ is an independent random variable uniformly distributed over all messages (both meaningful and meaningless).

Consider the ciphertext $C=E_K(M)$ for given K and M.

A **spurious key decipherment** (false solution) arises whenever encipherment under another key K′ could produce C for the same message M or for another meaningful message M′.

Approximating Unicity Distance

For every correct solution to a particular ciphertext, there are $2^{H(K)}-1$ remaining keys, each having the same probability of producing a spurious key decipherment.

Because each plaintext message is equally likely, the probability of getting a meaningful message (and so a false solution) is
$q = (2^{rN}/2^{RN}) = 2^{(r-R)N} = 2^{-DN}$

The expected number F of false solutions is
$F = (2^{H(K)}-1)q = (2^{H(K)}-1)2^{-DN} \cong 2^{H(K)-DN}$

Approximating Unicity Distance

The number of false solutions is sufficiently small to break the cipher when

$\log_2 F = H(K) - DN = 0$

and so $N = H(K)/D$ is taken as an approximation to the unicity distance, the amount of text necessary to break the cipher.

Note that in a theoretically unbreakable cipher the number of possible keys is as large as the number of messages of a given length N so that
$H(K) = \log_2(2^{RN}) = RN$ and
$H(K) - DN = (R-D)N = rN \neq 0.$

Approximating Unicity Distance

Example. Consider a simple substitution cipher with a shift of K positions, $0 \leq K \leq 25$. What is the unicity distance?

$N=4.6/3.2 = 1.5$ characters.

The estimate doesn't seem plausible: no substitution cipher can be broken with just one or two characters.

Usefulness of Unicity Distance

There are two reasons why the estimate is not very good.

- The estimate D=3.2 applies only to reasonable long messages.
- The cipher is not a good approximation to the random cipher model since most ciphertexts are not produced by meaningful messages (e.g., QQQQ) an so the decipherments are not uniformly distributed over the entire message space.

Usefulness of Unicity Distance

- The random cipher model gives a lower bound of the amount of ciphertext needed to break a cipher, a particular cipher will have a unicity distance at least H(K)/D.

- Note that the unicity distance gives the number of characters required to uniquely determine the key but it does not indicate the computational difficulty of finding it.

Usefulness of Unicity Distance

- A cipher may be computationally infeasible to break even if it is theoretically possible to break it with a small amount of ciphertext (e.g., AES).

- On the other hand, many substitution ciphers which use longer keys and have much greater unicity distance than AES are relatively simple to break when enough ciphertext is intercepted.

- The unicity distance N is inversely proportional to the redundancy D.

- As D approaches 0, an otherwise trivial cipher becomes unbreakable.

Usefulness of Unicity Distance

Example: Suppose M is a 6-digit integer enciphered 351972 using a Caser-type substitution cipher with key K, $0 \leq K \leq 9$, and that all possible 6-digit integers are equally likely. How much ciphertext is needed to break the cipher?

Answer: Such a cipher cannot be broken because there is no redundancy - no matter how much ciphertext is available.

Obstructing Cryptanalysis

Natural languages have an inherent redundancy which can be exploited to solve many ciphers by statistical analysis: frequency distribution of letters, ciphertext repetition, probable words.

Protecting against statistical analysis can be provided by removing some of the redundancy of the language before encryption, using data compression.

Obstructing Cryptanalysis

Confusion involves substitutions that make the relationship between the key and the ciphertext as complex as possible.

Diffusion involves transformations that dissipate the statistical properties of the plaintext across the ciphertext.

Many modern ciphers provide confusion and diffusion through complex enciphering transformations over large blocks of data.

Introduction to Cryptography

- Symmetric encryption, or conventional / secret-key / single-key:
  - sender and recipient share a common key
  - all classical encryption algorithms are secret-key
  - was only type prior to invention of public-key in 1970's

- Public-key encryption:
  - sender's and recipient's keys are neither the same nor easily derived from each other
  - has advantage of not having to exchange keys

- In what follows we will refer to symmetric encryption, unless stated otherwise

Introduction to Cryptography

- In terms of the type of encryption operations used, we distinguish between
  - Substitution ciphers
  - Transposition ciphers
  - Product ciphers

- In terms of the way in which plaintext is processed, we distinguish between
  - Block ciphers
  - Stream ciphers

Introduction to Cryptography

- In terms of the type of encryption operations used, we distinguish between
  - Substitution ciphers
  - Transposition ciphers
  - Product ciphers

- In terms of the way in which plaintext is processed, we distinguish between
  - Block ciphers
  - Stream ciphers

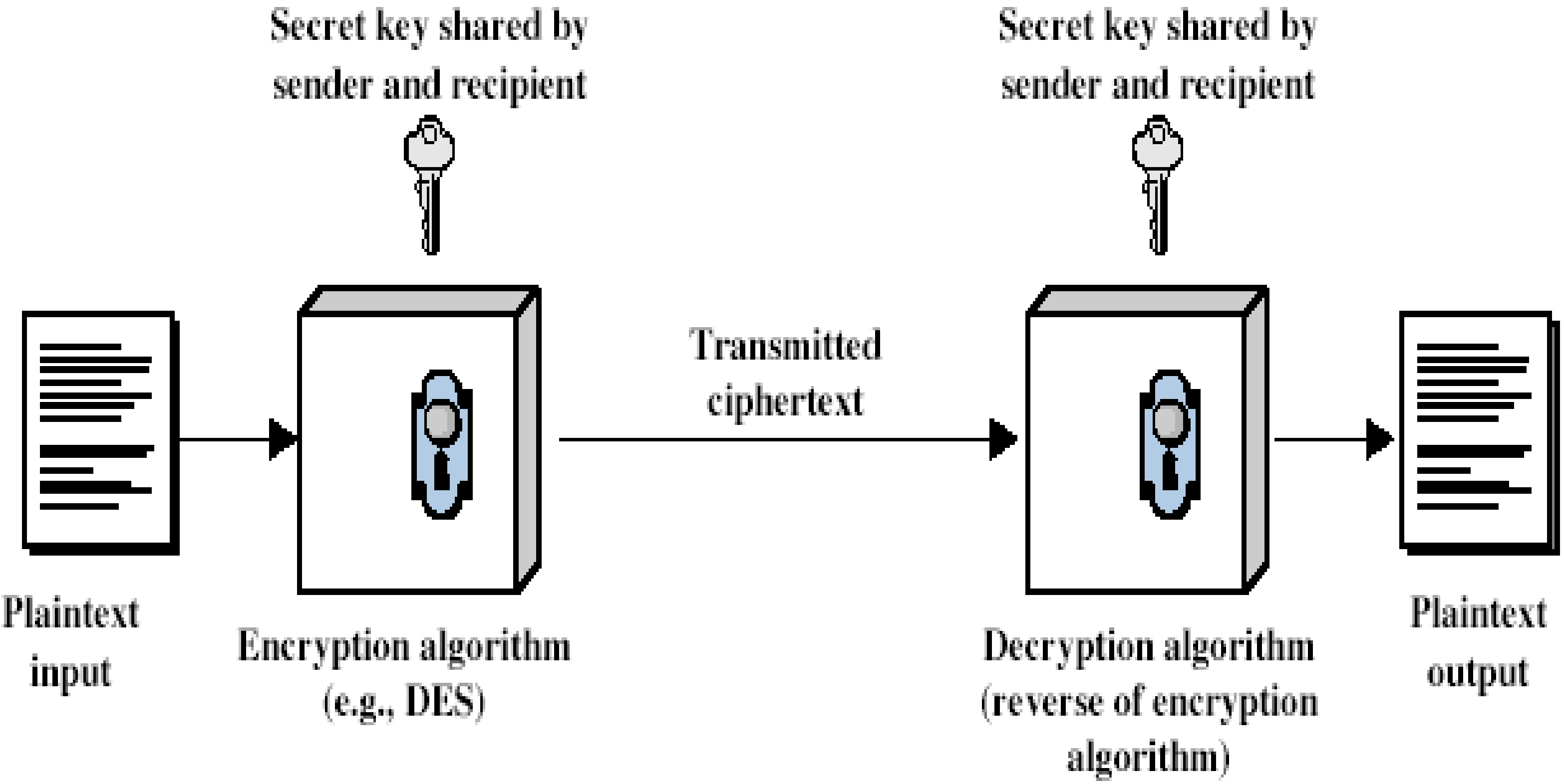
Basic Terminology

- **Plaintext** - the original message
- **Ciphertext** - the code (“encrypted”) message
- **Cipher** - algorithm for transforming plaintext to ciphertext
- **Key** - information used in cipher known only to sender/receiver

Basic Terminology

- **Enciphering (encrypting)** - converting plaintext to ciphertext
- **Deciphering (decrypting)** - recovering plaintext from ciphertext
- **Cryptography** - study of encryption principles/methods
- **Cryptanalysis (codebreaking)** - the study of principles/ methods of deciphering ciphertext *without* knowing the key
- **Cryptology** = Cryptography + Cryptanalysis

Symmetric Cipher Model


Secret key shared by sender and recipient
Secret key shared by sender and recipient
Transmitted ciphertext
Plaintext input
Encryption algorithm (e.g., DES)
Decryption algorithm (reverse of encryption algorithm)
Plaintext output

Requirements

- There are two requirements for secure use of symmetric encryption:
  - a strong encryption algorithm
  - a secret key known only to sender / receiver

    $Y = E_K(X)$

    $X = D_K(Y)$

- The security of an encryption system should only depend on the secrecy of the key and not the secrecy of the encryption algorithm.

- We need a secure channel to distribute keys.

Kerckhoffs' law

(Auguste Kerckhoffs, Professor of Linguistics and cryptographer, 1835 - 1903)

A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.

Kerckhoffs' laws

In 1883 Kerckhoffs published six principles of practical cipher design:

1. The system should be, if not theoretically unbreakable, unbreakable in practice.
2. Compromise of the system should not inconvenience the correspondents.
3. The key should be rememberable without notes and should be easily changeable.
4. The cryptograms should be transmittable by telegraph.
5. The apparatus or documents should be portable and operable by a single person.
6. The system should be easy, neither requiring knowledge of a long list of rules nor involving mental strain.

Kerckhoffs' law

Shannon's maxim: "The enemy knows the system."

Bruce Schneier:

> "Kerckhoffs' principle applies beyond codes and ciphers to security systems in general: every secret creates a potential failure point. Secrecy, in other words, is a prime cause of brittleness—and therefore something likely to make a system prone to catastrophic collapse. Conversely, openness provides ductility."

Security through Obscurity

Security through obscurity (security by obscurity [3]) uses secrecy of the encryption algorithm to ensure security.

Problems:

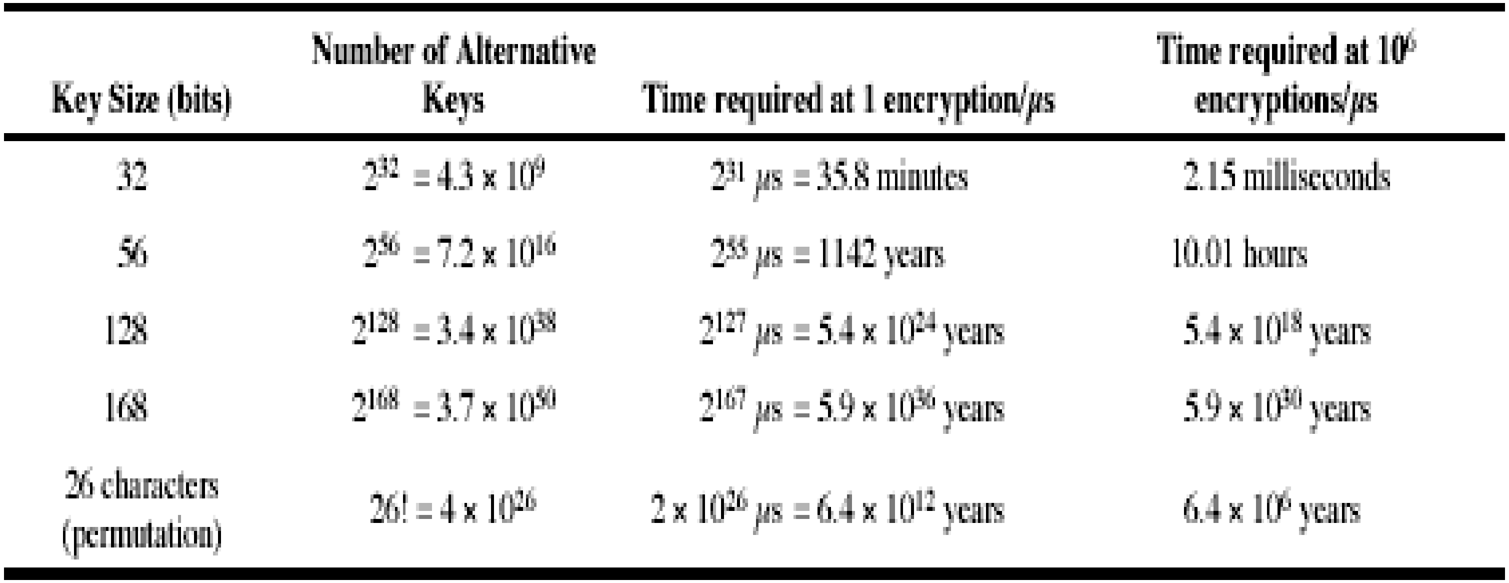
- Experience shows that secret algorithm designs are eventually disclosed either through reverse engineering or by leaked information. Thus if the system has weaknesses it cannot be subsequently used.
- The more secrets a system has, the less secure it is [3].
- If the algorithm is kept secret, the opportunities for security reviews and improvements are limited [3].

Types of Cryptanalytic Attacks

- **Ciphertext only**
  - only know algorithm/ciphertext
- **Known plaintext**
  - know/suspect plaintext and ciphertext
- **Chosen plaintext**
  - select plaintext and obtain ciphertext
- **Chosen ciphertext**
  - select ciphertext and obtain plaintext
- **Chosen text**
  - select either plaintext or ciphertext and obtain the other one

Brute Force Search

- It is always possible to simply try every key
- This is the most basic attack, proportional to key size.
- We are assuming that an intruder can recognise the plaintext.



Key Size (bits)	Number of Alternative Keys	Time required at 1 encryption/μs	Time required at 10^6 encryptions/μs
32	$2^{32} = 4.3 \times 10^9$	2^{31} μs = 35.8 minutes	2.15 milliseconds
56	$2^{56} = 7.2 \times 10^{16}$	2^{55} μs = 1142 years	10.01 hours
128	$2^{128} = 3.4 \times 10^{38}$	2^{127} μs = 5.4×10^{24} years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	2^{167} μs = 5.9×10^{36} years	5.9×10^{30} years
26 characters (permutation)	$26! = 4 \times 10^{26}$	2×10^{26} μs = 6.4×10^{12} years	6.4×10^6 years

Cryptography

Cryptography is the art (science, study) of writing in secret letters.

Secret writing:

1. Steganography
2. Cryptography

Steganography (concealment systems) hide the real message in covering messages which themselves look real, or attempt to hide even the existence of a message (e.g., invisible ink, microdots).

Cryptography does not conceal the existence of a message, only its meaning.

Codes

Cryptographic systems:

- code systems
- cipher system

Codes are mappings which are semantic in nature and which map letters, words, and/or entire messages into encoded text by means of a predefined table.

Advantage: by correctly designing a code, it is possible to make the encoded text appear to be a message of entirely different meaning.

Disadvantage: the need for a substitution table (or code-book) entry for every possible message severely restricts the types of messages which can be encoded.

Codes

For general computer systems, using coding techniques to achieve security is:

- too restrictive (usually impossible to predict types of messages)
- for general communication the code-book would have to be very large and kept in a very safe place - impractical for computer systems.

Ciphers are more flexible that codes.

Ciphers

Classical ciphers fall into one of the following categories:

- transposition ciphers, where the characters in the plaintext are simply rearranged
- substitution ciphers, where each character (or a group of characters) is substituted by another character (or a group of characters); substitution ciphers can be divided into:
  - monoalphabetic
  - homophonic
  - polyalphabetic
  - polygrams

More Definitions

Perfect Secrecy

- no matter how much computer power is available, the cipher cannot be broken since the ciphertext provides no information whatsoever about corresponding plaintext

Unconditional security

- no matter how much computer power is available, the cipher cannot be broken since the ciphertext provides insufficient information to uniquely determine the corresponding plaintext

Computational security

- given limited computing resources the cipher cannot be broken – eg, time needed for calculations is greater than age of universe

Next Week

1. Classical ciphers
    1. Transposition Ciphers and How to Break Them
    2. Substitution Ciphers and How to Break Them

- Text Chapter 2
- “Cryptography and Data Security” by D. Denning - Information theory

References

1. W. Stallings. “Cryptography and Network Security”, Pearson, global edition, 2016.

2. D. Denning. “Cryptography and Data Security”, Addison Wesley, 1982.