Cryptology

- Cryptology is the complete science of secure communications, formed from the Greek words kryptos, "hidden", and logos, "word".

- The science of cryptology is at least 2,000 years old. Modern advances in computer technology have made it possible to encipher huge amounts of data very securely in a reasonable time, while at the same time making it easier for the cryptanalyst to break cryptographic systems.

- It covers both Cryptography and Cryptanalysis (note that some misguided literature still equates Cryptology to magic).

- Cryptography is the science of the enciphering and deciphering of messages in secret code or cipher. This is done using various methods ("ciphers") to keep messages secret.

- Cryptanalysis is the science (and art) of recovering information from ciphers without knowledge of the key. It is the process of attacking ciphers, finding weaknesses, or even in the process, proving

Some Basic Definitions and Terminology

- **Plaintext**:
    - The source information to be secured.
- **Ciphertext**:
    - The encrypted form of the Plaintext.
- **Cipher:**
    - A map from a space of Plaintext to a space of Ciphertext.
- **Stream cipher:**
    - A cipher which acts on the plaintext one symbol at a time.
- **Block cipher**:
    - A cipher which acts on the plaintext in blocks of symbols.
- **Code**:
    - Not to be confused with ciphertext. An unvarying rule for replacing a piece of information with another object, not necessarily of the same sort.

- The ASCII code is a simple example. A code may consist of large number of words, phrases, letters, and syllables with codewords or codenumbers (also referred to as **codegroups**) that can be used to replace plaintext.

- Example:

codenumber	plaintext
121	country
304	money
690	without
691	within

- **Encryption**:
  - The process of converting the Plaintext into a Ciphertext.
- **Decryption**:
  - The process of converting the Ciphertext back into Plaintext.
- **Encryption/Decryption Key**:
  - The secret information known only to the transmitter and the receiver which is used to encrypt the Plaintext as well as to decrypt the ciphertext according to a regular algorithm or system.
  - These may or not be the same. They also need not be secret (public key encryption).
- **Substitution cipher**:
  - A stream cipher which acts on the plaintext by making a substitution of the characters with elements of a new alphabet or by a permutation of the characters in the plaintext alphabet.
  - The units of plaintext to be replaced with ciphertext may be single letters (the most common), pairs of letters, triplets of letters, or combinations of them.
  - Decryption is carried out by performing an inverse substitution.
  - There are a number of different types of substitution ciphers: **simple substitution cipher**; **monoalphabetic cipher, polyalphabetic cipher** and **the one-time pad**.

- **Simple substitution cipher:**
    - This is the simplest form of encryption where each letter in the plaintext is replaced by a letter some fixed number of positions in the alphabet.
    - The simplest form is the Caesar cipher (named after Julius Caesar who used it for his secret communications). For example, with a shift of 3, J would be replaced by M; C would become F, and so on.
    - As with all single alphabet substitution ciphers, the Caesar cipher is easily broken and in modern practice offers absolutely no communication security.
    - This cipher can be improved by using a scrambled (random) alphabet instead of a simple offset. This is referred to as a mixed or deranged alphabet.

- **Monoalphabetic Substitution**:
    - A method of encryption whereby a letter in the Plaintext is always replaced by the same letter in the Ciphertext.

- **Polyalphabetic Substitution**:
    - A method of encryption whereby a letter in the Plaintext is not always replaced by the same letter in the Ciphertext.
    - It is implemented using multiple substitution ciphers. For example, there might be four different simple substitution ciphers used, the particular one used changes with the position of each character in the plaintext.

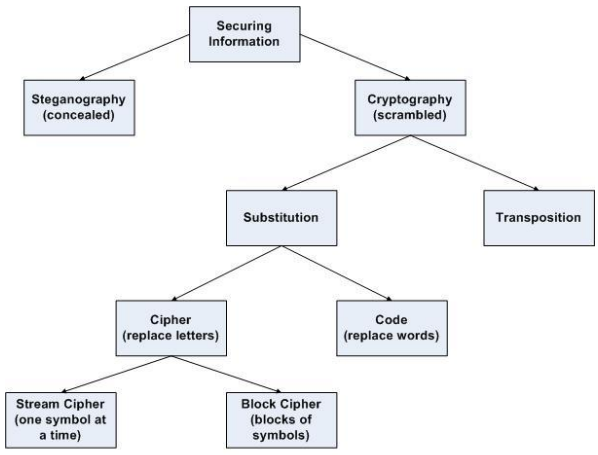
- **The one-time pad:**
    - This is a special case of the substitution cipher, invented near the end of WWI and was mathematically proven unbreakable by Claude Shannon during WWII (published in the late 1940s).
    - Typically, the plaintext letter is combined (not substituted) in some manner (e.g., XOR) with the key character at that position.
    - The one-time pad is not very impractical as it requires that the key being used in the encryption process be as long as the plaintext, in fact **random**, used once and only once, and kept entirely secret from all except the sender and intended receiver.
    - If these conditions are violated, even marginally, the one-time pad is no longer unbreakable.

- **Transposition**:
    - A method of encryption whereby letters or words in the Plaintext are rearranged based on a rule or permutation. The letters are not actually but transposed relative to each other.
    - In a transposition cipher, the units of the plaintext are rearranged in a different and usually quite complex order, but the units themselves are left unchanged.
    - By contrast, in a substitution cipher, the units of the plaintext are retained in the same sequence in the ciphertext, but the units themselves are altered.

- **Stream Ciphers**:
    - The data is processed in very small increments like a stream of bits (also referred to as a bit-manipulation cipher).
    - They operate on one bit or byte at a time and use a relatively simple operation to combine this plaintext stream with a pseudorandom **keystream**. The operation that combines these two streams is typically **Exclusive-Or (XOR).**
    - Stream ciphers typically execute at a higher speed than block ciphers and have lower hardware complexity. However, stream ciphers can be susceptible to serious security problems if used incorrectly, especially if the same starting state (seed) is used twice.

- **Block Ciphers**:
    - In these algorithms, data is processed in blocks (64 or 128-bit block size for example). The same key is typically used for each block.
    - It simplifies things if the key size is the same as the block size, but it need not be. A part of the cipher called the key expansion algorithm makes the key "fit" the block size.
    - Block ciphers are important basic building blocks in the design of many cryptographic protocols, and are widely used to implement encryption of bulk data.
    - The publication of the DES cipher by the U.S. National Bureau of Standards (now National Institute of Standards and Technology, NIST) in 1977 was a seminal block cipher implementation.
    - The DES cipher is fundamental in the study and understanding of modern block cipher designs.
    - In fact, both **Differential** and **Linear Cryptanalysis** arose out of studies on the DES design.

- The following diagram summarizes the main branches of the science of securing data:



The Science of Securing Information

Basic Notation

- The science of cryptology uses a standard notation and it is important to become familiar with this notation.

- Plaintext is denoted by **M**, for message or **P**, for plaintext. This is the original readable message (written in some standard language, like English, French, Japanese, Hindi, Swahili, etc.). As far as a computer is concerned, M is binary data.

- Plaintext is denoted by **C**. This is the output of some encryption scheme, and is not readable by ordinary means. This is also binary data and not necessarily the same size as M.

- The Encryption function **E**, operates on M to produce C. In mathematical notation, this is written as:

  E(M) = C

- During the reverse process, the decryption function D operates on C to produce M:

  D(C) = M

- Thus in the following identities hold true:

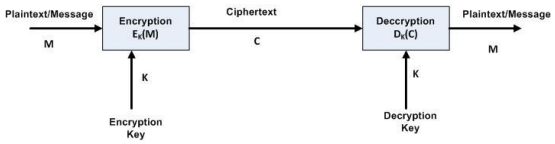
  D(E(M)) = M and E(D(C)) = C

Cryptographic Algorithms (Ciphers)

- Cryptographic algorithms are mathematical functions used for encryption and decryption.
- Modern ciphers use a key(s), denoted by **K**. The range of possible values for the key is called the **keyspace**.
- Both the encryption and decryption operations use these keys, thus the aforementioned functions now become:

  $\mathbf{E_K(M) = C}$
  $\mathbf{D_k(C) = M}$
  $\mathbf{D_K(E_K(M)) = M}$

- The security of these algorithms is dependent entirely on the keys and not on the details of the algorithm. In other words as long as the keys remain secret the message will be unreadable even if the algorithm is known.
- Such algorithms are known as Symmetrical algorithms where the encryption key can be calculated from the decryption key and vice versa. In other words both keys are the same.
- These algorithms are also referred to secret algorithms, which require that the sender and receiver both agree upon the key before they initiate secret communications.
- The diagram below depicts this model:

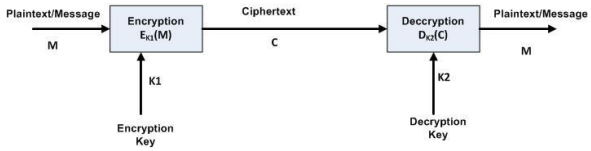


Cryptography using the same key

- Some algorithms use different keys for encryption and decryption, in which case the functions can be stated as follows:

  $\mathbf{E_{K1}(M) = C}$
  $\mathbf{D_{k2}(C) = M}$
  $\mathbf{D_{K2}(E_{K1}(M)) = M}$

- The diagram below depicts this model:



Cryptography using different keys

Public Key Algorithms

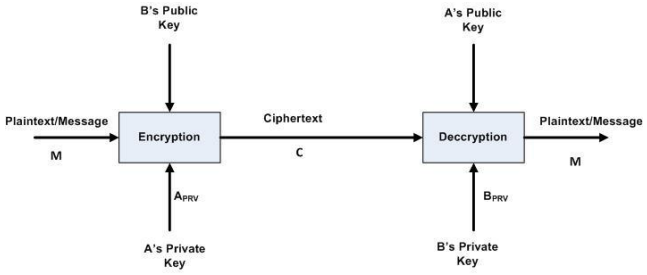
- The main problem with secret keys is exchanging them over a public medium such as the Internet or a network, while preventing them from falling into the wrong hands.

- The solution lies in using an asymmetric algorithm in which the encryption and decryption processes use different keys. In addition the decryption key cannot be (in any reasonable amount of time) be calculated from the encryption key.

- These algorithms are known as “public-key” algorithms because the encryption can be made public. A public key is made freely available to anyone who might want to send you a message. A second, private key is kept secret, so that only you know it.

- These **key pairs** are mathematically generated and are related to each other in such a way that encrypting text with one (called the public key) requires decryption with the other key (called the private key).
- Any message (text, binary files, or documents) that are encrypted by using the public key can only be decrypted by applying the same algorithm, but by using the matching private key. Any message that is encrypted by using the private key can only be decrypted by using the matching public key.

- The diagram below illustrates this process:



Public Key Cryptography

Hashing Algorithms

- A **hash** is a short piece of data that represents a much larger piece of data (like a file). Hash functions take the original data in and produce a hash of it. This is a one-way process, meaning that this process cannot be performed in reverse.

- Because of the difference in the amount of input data and the size of the hash, many data sets will hash to the same value. The trick is that the hashing algorithm should make it very hard/impossible to find these hash **collisions**.

- A very good example of a hashing application is the distribution of public-domain software. When users download the piece of software, they will also be provided with the hash of the original file contents.

- They can then calculate the hash by running it through the same algorithm. If the has is the same as the one downloaded, they can be sure (assuming the hashing algorithm is strong) that the data has not been tampered with. It would be computationally infeasible to modify the file in any targeted way and still have it produce the same hash.

Cryptanalysis

- Cryptanalysis is the science of recovering the plaintext of a message without access to the key. Successful cryptanalysis may recover the plaintext or the key.

- To attack a cipher is to attempt unauthorized reading of plaintext, or to attempt unauthorized transmission of ciphertext.
- There are four general types of cryptanalysis attacks, each one is premised on the attacker having complete knowledge of the encryption algorithm used:

  - **Ciphertext-only attack:**
    - The cryptanalyst has the ciphertext of several messages, all of which have been encrypted using the same encryption algorithm.
    - The objective here is to recover the plaintext of as many messages as possible, or to deduce the key (or keys) used to encrypt the messages.
    - This can be can be described as follows:

      Given: $\mathbf{C_1 = E_K(P_1), C_2 = E_K(P_2),\ldots\ldots C_n = E_K(P_n)}$
      Deduce: Either $\mathbf{P_1, P_2,\ldots P_n}$;
      Or an algorithm to infer $\mathbf{P_{n+1}}$ from $\mathbf{C_{n+1} = E_K(P_{n+1})}$

  - **Known-plaintext attack**:
    - The cryptanalyst has the ciphertext of several messages, and also all of the corresponding plaintext messages.
    - The objective here is to deduce the key (or keys) used to encrypt the messages, or an algorithm to decrypt any subsequent new messages encrypted with the same key.
    - This can be can be described as follows:

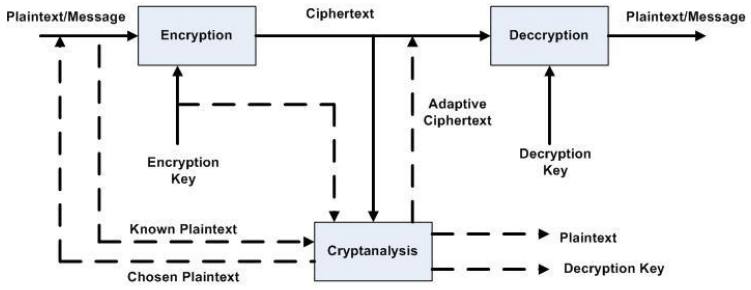
      Given: $\mathbf{P_1, C_1 = E_K(P_1), P_2, C_2 = E_K(P_2),\ldots\ldots P_n, C_n = E_K(P_n)}$
      Deduce: Either **K**;
      Or an algorithm to infer $\mathbf{P_{n+1}}$ from $\mathbf{C_{n+1} = E_K(P_{n+1})}$

  - **Chosen-plaintext attack**:
    - The cryptanalyst has access to not only the ciphertext of several messages, and also all of the corresponding plaintext messages, but can also chose the plaintext that gets encrypted.
    - This is a much more powerful attack because the cryptanalyst can choose specific plaintext blocks to encrypt , ones that might yield more information about the key.
    - The objective here is to deduce the key (or keys) used to encrypt the messages, or an algorithm to decrypt any subsequent new messages encrypted with the same key.
    - This can be can be described as follows:

      Given: $\mathbf{P_1, C_1 = E_K(P_1), P_2, C_2 = E_K(P_2),\ldots\ldots P_n, C_n = E_K(P_n),}$ where the cryptanalyst gets to choose $\mathbf{P_1, P_2,\ldots P_n}$
      Deduce: Either **K**;
      Or an algorithm to infer $\mathbf{P_{n+1}}$ from $\mathbf{C_{n+1} = E_K(P_{n+1})}$

- **Adaptive-ciphertext attack**:
    - This is a special case of a chosen-plaintext attack. Not only can the cryptanalyst choose specific plaintext blocks to encrypt, but he can also modify his choice based on the results of previous encryption.
    - Using this attack the cryptanalyst might just be able to choose one large block of plaintext to be encrypted, then subsequently use smaller blocks based on the results of the previous results and so on.

- The following diagram summarizes all of the processes described so far:



Cryptography and Cryptanalysis