The desire to transmit mssages securely is not new. For centuries, people have had a need to keep their communications private. Today, digital communications systems, particularly those related to the Internet, are used to carry vast amounts of sensitive data. Sending credit card information to a web site in an e-commerce transaction or exchanging confidential trade secrets by e-mail are typical examples.

The field of *cryptography* deals with the techniques for conveying information securely. The goal is to allow the intended recipients of a message to receive the message properly while preventing eavesdroppers from understanding the message. The message in its original form is called *plaintext*. The transmitter in a secure system will encrypt the plaintext in order to hide its meaning. This reversible mathematical process produces an encrypted output called *ciphertext*. The algorithm used to encrypt the message is a *cipher*. *Cryptanalysis* is the science of breaking ciphers, and *cryptanalysts* try to defeat the security of cryptographic systems.

A ciphertext can be transmitted openly across a communications channel. Because of its encrypted nature, eavesdroppers who may have access to the ciphertext will ideally be unable to uncover the message's meaning. Only the intended recipient can decrypt the message to recover the plaintext for interpretation. These processes are shown in Fig. 1.

## Classifying ciphers

Ciphers can be classified using several criteria. According to one criterion, two important types of ciphers exist: symmetric key and asymmetric key.

In symmetric key ciphers, the same key is used for both encryption and decryption. A major problem with such a system is that the sender and receiver must know the key prior to transmission. This requirement makes such a system difficult to use in practice. The key cannot be openly transmitted since that would compromise the system's security. One possibility is for the two parties to meet and exchange the keys prior to transmitting their messages. However, this exchange becomes difficult when many parties are involved in a communications network.

An asymmetric key cipher uses different keys for encryption and decryption. These two keys are mathematically related, but it is very difficult to obtain one from the other. The key used for encryption is called the public key and the key used for decryption is called the private key. The public key can be made available without compromising the security of the system. The corresponding private key, however, must not be revealed to any party.

## Classical ciphers

Developed and used before the Computer Age, these ciphers are not secure against today's cryptanalysis. They are included for their historical importance and educational value.

## English stats

A plaintext message written in English (or any other language) has certain statistical characteristics. The letter *E* is the most frequent in the English language. The next most frequent is *T*, followed by *O*, *A*, *N*, *I*, *R*, *S*, *H* and so forth. *Z* is the least frequent. All ciphers modify the plaintext to change its statistics. A cipher is considered to be "weak" if it generates ciphertext that still contains significant statistical information about the original plaintext. Cryptanalysis of classical ciphers is made possible because of the redundancy in the linguistic structure of natural languages.

## Simple substitutions

Simple (monoalphabetic) substitution ciphers replace each letter in the plaintext with another letter to form the ciphertext. This class of ciphers is easy to implement and use. They also are not difficult to break and, thus, do not offer much security.

An eavesdropper can decrypt a ciphertext by performing frequency analysis on the letters in the ciphertext. Because of the substitution, the letter frequencies will be different than the frequencies for normal English text. The eavesdropper can still exploit this frequency information to break the cipher. If, for example, *Z* is the most frequent letter in the ciphertext, then it was probably substituted in for *E*. The next most frequent letter in the ciphertext may correspond to a *T* or an *O*, and so on. By trial and error, the entire plaintext message can be revealed. Simple substitution ciphers are also called monoalphabetic substitution ciphers because they define a mapping from the plaintext alphabet to a ciphertext alphabet.

Although simple substitution ciphers are not actually used in today's encryption systems, powerful modern ciphers do use substitution in combination with other operations (transposition, Boolean algebra, modular arithmetic, etc.). This is a very important design criterion that results in an algorithm more secure than its components.

## Caesar cipher

A simple substitution cipher of historical interest is the Caesar cipher. The cipher gets its name because Julius Caesar (100 B.C.E. - 44 B.C.E.) used it to send secret messages. The ciphertext is formed by replacing each letter in the plaintext by the letter three positions to the right in the alphabet. This shift is performed modulo 26. Thus, the plaintext letter *A* becomes *D*, *B* becomes *E* and *Z* becomes *C*.

Breaking this cipher is easy


since the plaintext's statistical information is contained in the ciphertext; however, keep in mind that the letter frequencies are also shifted to the right by three. Instead of *E* being the most frequent letter, the letter *H* will be the most frequent in the ciphertext. Frequency analysis quickly reveals that the Caesar cipher is being used. Thus, it is a simple matter to replace the ciphertext letters in order to uncover the plaintext message.

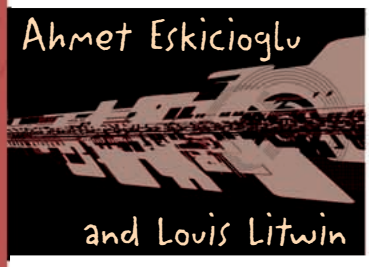
The amount of the shift $K = 3$ is defined to be the key for the Caesar cipher. Shifts by an amount other than three can also be used. Nevertheless, there are only 25 possible shifts for the English alphabet, and it is easy to try all possible combinations if necessary.

*An example of the Caesar cipher*. The example in Table 1 is an application of the Caesar cipher. The most frequent letters in the ciphertext are simply shifted versions of the most frequent letters in the plaintext.

The ciphertext alphabet used in the Caesar cipher is very orderly since it is only a shifted version of the plaintext alphabet. Some simple substitution ciphers use a scrambled alphabet that has no apparent order. These ciphers are slightly harder to break than the Caesar cipher. This is because there is not such an obvious pattern to the ciphertext alphabet. For example, *A* might be substituted with *Q*, *B* with *D* and *C* with *R*. The letters in the English alphabet can be rearranged in 26! (over $4 \times 10^{26}$) different ways, creating a large number of possible ciphertext alphabets. It would, of course, take a significantly long time to try all possible combinations. However, all simple substitution ciphers can be easily broken using frequency analysis.

## Polyalphabetic subs

Polyalphabetic substitution ciphers use multiple alphabets to conceal the single letter frequency distribution of the plaintext letters in the ciphertext. In its simplest




form, they are based on a period *d* that determines the number of the alphabets. Using the alphabets sequentially results in a given plaintext character being encrypted to different ciphertext characters. For a monoalphabetic cipher, *d* is equal to 1.

*Example of a polyalphabetic substitution cipher*. The simple Vigenère cipher is representative of polyalphabetic substitution ciphers with a period. The letter $k_i$ in the key $K = k_1 \ldots k_d$ determines the amount of shift in the *i*th alphabet. Periodic substitution ciphers can be cryptanalyzed in two steps. First, the period *d* is estimated. The Kasiski method and the Index of Coincidence are two useful tools for this purpose. The work is then reduced to the cryptanalysis of a set of monoalphabetic substitution ciphers.

## Transposition ciphers

Transposition ciphers are a different family of ciphers. Frequency analysis does not provide any useful information in an attempt to obtain the plaintext. In fact, transposition ciphers produce ciphertext that has the exact same letter frequencies as the original plaintext. The reason is that they work by rearranging, or transposing, the letters in the plaintext. Thus, the ciphertext contains the same letters as the plaintext, but the order of the letters is changed.

A common method of transposition is to insert the plaintext into a matrix in some known way (e.g., by inserting the text by rows), and forming the ciphertext by reading the letters out in another known way (e.g., by reading them out the columns). This information determines the key for the cipher. More complicated transposition ciphers can be formed building upon this basic idea.

*An example of a transposition cipher*. The example in Table 2 is for an application of a transposition cipher. The cipher uses the transposition matrix shown in Table 3. Note that the frequency analysis gives the same results for both the plaintext and the ciphertext. The ciphertext, in this example, was formed by writing the plaintext into the rows of a $5 \times 4$ matrix. We then read the text out by the columns.

Transposition ciphers can be broken with restoring the original order of the letters. Column and row rearrangements, and frequency distributions of digrams (two-letter sequences) and trigrams (three-letter sequences) are commonly used in cryptanalysis of transposition ciphers.

## Modern ciphers

With high-speed digital computing machines, classical ciphers have become inadequate for providing information security. After World War II, a need emerged for stronger ciphers that could be used for non-military applications. Modern ciphers have been designed as an attempt to resist cryptanalytic attacks of today's powerful computer systems.

## DES

The Data Encryption Standard (DES), the well-known symmetric key cipher, was developed due to efforts initiated by the National Security Agency (NSA). In their public request for proposals, where a set of design criteria was specified, the NSA argued that the security of the algorithm must reside in the key. In 1977, DES was adopted as a federal standard for use in commercial and unclassified U.S. government applications. In later years, both hardware and software implementations became widely available. They have been used in many sectors of industry including banking.

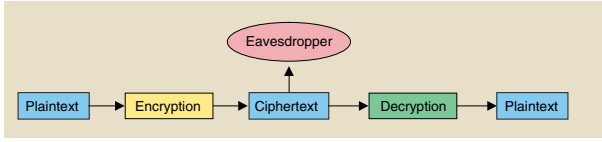



*Fig. 1 Block diagram of a cryptographic system*

DES is a product cipher; i.e., it is a composition of substitutions and transpositions. Like all other modern ciphers, the algorithm is published in full detail. There are two major arguments for not keeping a cryptographic algorithm secret: 1) It is very difficult to keep the algorithm secret, especially if it is to be used in a standard. 2) Publication leads to open discussions and cryptanalysis to evaluate the real strengths and weaknesses.

DES operates on 64-bit blocks of plaintext to produce 64-bit ciphertext blocks. The length of the encryption key is 56 bits. Since DES is a symmetric cipher, this key is also used for decryption. DES keys are generated as 64-bit numbers, but in each key every eighth bit is used for error (parity) checking.

A summary of the DES algorithm is shown in Fig. 2. The output of the Initial Permutation (IP) is divided into two 32-bit halves $L_0$ and $R_0$. After 16 "rounds" of identical operations, the inverse permutation $IP^{-1}$ gives the ciphertext. The subkeys $K_1, K_2, \ldots, K_{16}$ are derived from the 56-bit encryption key. A combination of substitution and transposition operations define the function *f*.

The DES key's length and some other design aspects have been the subject of controversy. After more than 20 years of use, the original key length is no longer sufficient if high security is needed. Several DES variations have been proposed to increase the robustness of the algorithm. One popular implementation is triple DES (TDES) which uses three different keys (Fig. 3).

In the last few years, collaborative efforts were undertaken to develop a new algorithm, Advanced Encryption Standard (AES), to replace DES. A worldwide competition initiated by the National

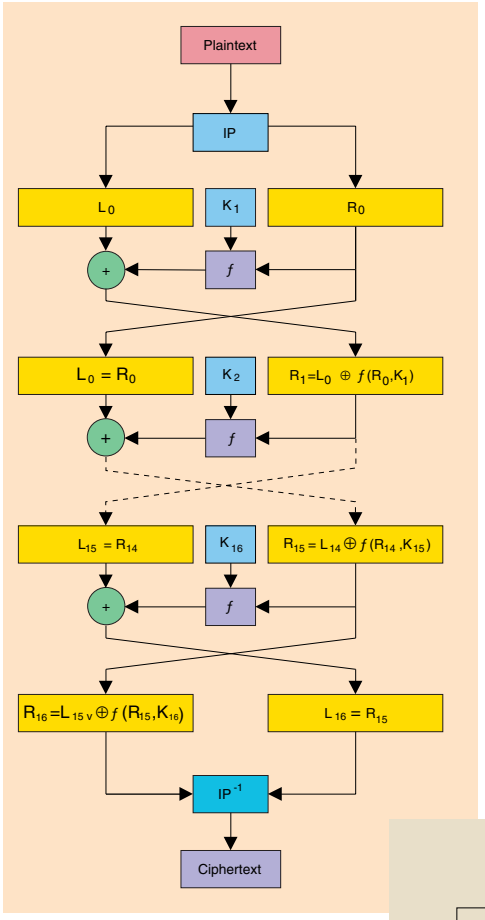



*Fig. 2 DES encryption*

Institute of Standards and Technology (NIST) in 1997 to develop a Federal Information Processing Standard (FIPS) ended last Fall with the announcement of Rijndael as the winner. The other AES candidate algorithms were MARS, RC6, Rijndael, Serpent, and Twofish. After a public comment period and subsequent revision, the NIST will submit the algorithm to the Secretary of Commerce for adoption as an official federal standard. It is expected that the process will be complete by this Spring (2001).

**Table 1 Example of the Caesar cipher**

| | Message | Five most frequent letters |
|---|---|---|
| Plaintext | THISISACAESARCIPHER | S, I, A, R, H |
| Ciphertext | WKLVLVDFDHVDUFLSKHU | V, L, D, U, K |

**Table 2 Example of a transposition cipher**

| | Message | Five most frequent letters |
|---|---|---|
| Plaintext | ATRANSPOSITIONCIPHER | I, T, S, R, P |
| Ciphertext | ANSOPTSINHRPTCEAOIIR | I, T, S, R, P |

## One-time pads

An encryption scheme that is perfectly secure actually exists (at least in theory). It is known as a one-time pad. The scheme gets its name from a pad that is used only once. The pad consists of a non-repeating random sequence of letters with a length equal to the plaintext length. Both the sender and the receiver will have a copy of the pad. Encryption works by adding each letter of the plaintext with the next letter on the pad. This addition is performed modulo 26 (so that the result is also a letter in the alphabet).

The decryption process is just the reverse. It is performed by adding the letters of the ciphertext with the letters on the pad modulo 26. If the pad is generated in a truly random fashion, the cipher cannot be broken as there is no pattern or statistics to exploit. Each letter of the alphabet has an equal probability of occurring in the ciphertext. However, the scheme is very cumbersome to implement. The pad can only be used once. One-time pads are best suited for sending low bit rate data that requires absolute security.

## Key agreement

As mentioned earlier, the strength of a modern cipher lies in its key, since the algorithm itself is typically published and, thus, well known. Because of the importance of cipher key secrecy, key management is a critical issue in secure communications. Several schemes have been developed so that communicating parties can establish a shared key. They are based either on symmetric or public key ciphers.

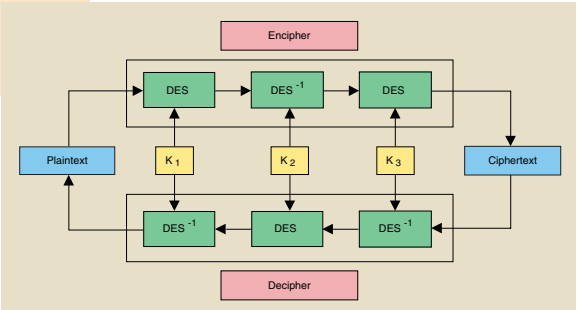



*Fig. 3 Triple-DES encryption and decryption*

*Diffie-Hellman public key exchange algorithm*. Diffie and Hellman developed the first, practical key exchange algorithm. Two parties, Alice and Bob, agree on numbers $g$ and $n$, which satisfy certain mathematical conditions. These numbers do not need to be kept secret and can be posted publicly.

1. Alice randomly selects a large number $x$ as her private key, and sends Bob her public key $X = g^x \bmod n$.

2. Bob randomly selects a large number $y$ as his private key, and sends Alice his public key $Y = g^y \bmod n$.

3. Alice computes $k_A = Y^x \bmod n$.

4. Bob computes $k_B = X^y \bmod n$.

Note that $k_A = (g^y)^x \bmod n$ and $k_B = (g^x)^y \bmod n$. Thus, $k_A = k_B = g^{xy} \bmod n$. This value becomes the shared key that will be used to encrypt and decrypt the message. Although $g$, $n$, $X$ and $Y$ are publicly available, it is very difficult to determine the private keys $x$ and $y$ based on these values.

## Conclusions

Early classical ciphers' substitution and transposition operations form the building blocks for today's powerful ciphers such as DES. Key agreement schemes allow communicating parties to establish a shared cipher key. The interested reader can learn more about this fascinating subject by starting with the references listed below.

## Read more about it


• D. Denning, *Cryptography and Data Security*, Addison-Wesley Publishing Company, 1983.

• B. Schneier, *Applied Cryptography*, John Wiley & Sons, Inc., 1996.

• A. J. Menezes, P. C. Van Oorschot, and S. A. Van Stone, *Handbook of Applied Cryptography*, CRC Press, 1997.

•AES Website: <http://csrc.nist.gov/encryption/aes/>


## About the authors


Ahmet M. Eskicioglu received his BS from the Middle East Technical University, Ankara, Turkey, and his MS and PhD from the University of Manchester Institute of Science and Technology (UMIST), England. Dr. Eskicioglu is a Principal Member Technical Staff with Thomson Multimedia Corporate Research, and is working on conditional access and copy protection projects. He has participated in the development of several national and international security standards.

Louis Litwin is a Member of the Technical Staff with Thomson Multimedia Corporate Research where he is working on wireless digital home networking technology.


**Table 3 Matrix used to create transposed ciphertext**

| | | | |
|---|---|---|---|
| A | T | R | A |
| N | S | P | O |
| S | I | T | I |
| O | N | C | I |
| P | H | E | R |