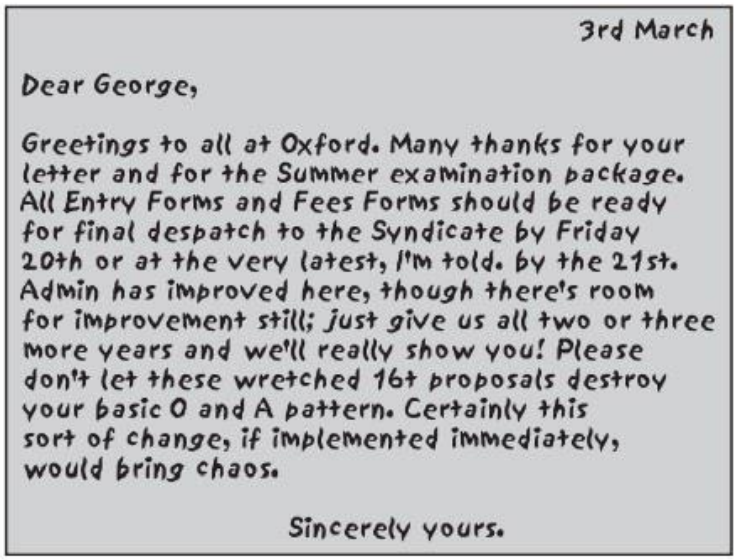
STEGANOGRAPHY

A plaintext message may be hidden in one of two ways. The methods of **steganography** conceal the existence of the message, whereas the methods of cryp-tography render the message unintelligible to outsiders by various transformations of the text.

A simple form of steganography, but one that is time-consuming to construct, is one in which an arrangement of words or letters within an apparently innocuous text spells out the real message. For example, the sequence of first letters of each word of the overall message spells out the hidden message. Figure 2.9 shows an example in which a subset of the words of the overall message is used to convey the hidden message. See if you can decipher this; it's not too hard.



3rd March

Dear George,

Greetings to all at Oxford. Many thanks for your letter and for the Summer examination package. All Entry Forms and Fees Forms should be ready for final despatch to the Syndicate by Friday 20th or at the very latest, I'm told. by the 21st. Admin has improved here, though there's room for improvement still; just give us all two or three more years and we'll really show you! Please don't let these wretched 16t proposals destroy your basic O and A pattern. Certainly this sort of change, if implemented immediately, would bring chaos.

Sincerely yours.

Figure 2.9 A Puzzle for Inspector Morse
(From The Silent World of Nicholas Quinn, *by Colin Dexter)*

Reference :William Stallings, Cryptography and Network Security: Principles and Practice, PHI 3rd Edition, 2006

Various other techniques have been used historically; some examples are the following:

• **Character marking:** Selected letters of printed or typewritten text are over-written in pencil. The marks are ordinarily not visible unless the paper is held at an angle to bright light.

• **Invisible ink:** A number of substances can be used for writing but leave novisible trace until heat or some chemical is applied to the paper.

• **Pin punctures:** Small pin punctures on selected letters are ordinarily not visible unless the paper is held up in front of a light.

• **Typewriter correction ribbon:** Used between lines typed with a black ribbon, theresults of typing with the correction tape are visible only under a strong light.

Although these techniques may seem archaic, they have contemporary equivalents. Hiding a message by using the least significant bits of frames on a CD. For example, the Kodak Photo CD format's maximum resolution is 2048 _ 3072 pixels, with each pixel containing 24 bits of RGB color information.

The least significant bit of each 24-bit pixel can be changed without greatly affecting the quality of the image. The result is that you can hide a 2.3-megabyte message in a single digital snapshot. There are now a number of software packages available that take this type of approach to steganography.

Steganography has a number of drawbacks when compared to encryption. It requires a lot of overhead to hide a relatively few bits of information, although using a scheme like that proposed in the preceding paragraph may make it more effective.Also, once the system is discovered, it becomes virtually worthless. This problem, too, can be overcome if the insertion method depends on some sort of key. Alternatively, a message can be first encrypted and then hidden using steganography.

The advantage of steganography is that it can be employed by parties who have something to lose should the fact of their secret communication (not necessarily the content) be discovered. Encryption flags traffic as important or secret or may identify the sender or receiver as someone with something to hide.

Foundations of modern cryptography

Modern cryptography is the cornerstone of computer and communications security. Its foundation is based on various concepts of mathematics such as number theory, computational-complexity theory, and probability theory.

Characteristics of Modern Cryptography

There are three major characteristics that separate modern cryptography from the classical approach.

Classic Cryptography	Modern Cryptography
It manipulates traditional characters, i.e., letters and digits directly.	It operates on binary bit sequences.
It is mainly based on 'security through obscurity'. The techniques employed for coding were kept secret and only the parties involved in communication knew about them.	It relies on publicly known mathematical algorithms for coding the information. Secrecy is obtained through a secrete key which is used as the seed for the algorithms. The computational difficulty of algorithms, absence of secret key, etc., make it impossible for an attacker to obtain the original information even if he knows the algorithm used for coding.
It requires the entire cryptosystem for communicating confidentially.	Modern cryptography requires parties interested in secure communication to possess the secret key only.

Perfect Security

- A cipher system is said to offer perfect secrecy if, on seeing the ciphertext the interceptor gets **no extra information**about the plaintext than he had before the ciphertext was observed.
- In a cipher system with perfect secrecy the interceptor is "forced" to guess the plaintext.

- An encryption scheme satisfies perfect secrecy if for all messages m1, m2 in message space M and all ciphertexts c ∈ C, we have

where both probabilities are taken over the choice of K in K and over the coin tosses of the (possibly) probabilistic algorithm Enc.

- Intuitively, we might want to define perfect security of an encryption scheme as follows: Given a ciphertext all messages are equally likely.
- This can be formulated as: For all m(0) , m(1) ∈ M and c ∈ C we have:

  Pr[M = m (0) |C = c] = Pr[M = m (1) |C = c]

- The probability here is over the randomness used in the Gen and Enc algorithms and the probability distribution over the message space.

Definition (One: Perfect Security)

- We want the ciphertext to provide no additional information about the message
- Definition (One: Perfect Security)

  For all m ∈ M and c ∈ C, we have:

  Pr[M = m|C = c] = Pr[M = m]

- Here we are assuming that c ∈ C has Pr[C = c] > 0. Everywhere this assumption will be implicit

Definition (Two: Perfect Security)

- We want to say that the probability to generate a ciphertext given a message is independent of the message
- Definition (Two: Perfect Security)

  For all m ∈ M and c ∈ C we have:

  Pr[C = c|M = m] = Pr[C = c]

Definition (Three: Perfect Security)

- We want to say that the probability of generating a ciphertext given as message m(0) , is same as the probability of generating that ciphertext given any other different message m(1)
- Definition (Three: Perfect Security)

For any messages m(0) , m(1) ∈ M and c ∈ C we have:

$$Pr[C = c|M = m(0)] = Pr[C = c|M = m(1)]$$

Shannon's Original Definition of Secrecy

- Shannon defines perfect secrecy for secret-key systems and shows that they exist.
- A secret-key cipher obtains perfect secrecy if for all plaintexts *x* and all ciphertexts *y* it holds that $Pr(x) = Pr(x|y)$.
- In other words, a ciphertext y gives no information about the plaintext

Information theory

- **Information theory** studies the quantification, storage, and communication of information.
- A key measure in information theory is entropy. Entropy quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process.
- For example, identifying the outcome of a fair coin flip (with two equally likely outcomes) provides less information (lower entropy) than specifying the outcome from a roll of a die (with six equally likely outcomes).
- Some other important measures in information theory are mutual information, channel capacity, error exponents, and relative entropy.

Quantities of information

- Information theory is based on probability theory and statistics.

- Information theory often concerns itself with measures of information of the distributions associated with random variables.
- Important quantities of information are entropy, a measure of information in a single random variable, and mutual information, a measure of information in common between two random variables.

A common unit of information is the bit, based on the binary logarithm

Entropy of an information source

- Based on the probability mass function of each source symbol to be communicated, the Shannon entropy *H*, in units of bits (per symbol), is given by

$$H = -\sum_i p_i \log_2(p_i)$$

- where p_i is the probability of occurrence of the *i*-th possible value of the source symbol.
- This equation gives the entropy in the units of "bits" (per symbol) because it uses a logarithm of base 2, and this base-2 measure of entropy has sometimes been called the shannon in his honor.
- If one transmits 1000 bits (0s and 1s), and the value of each of these bits is known to the receiver (has a specific value with certainty) ahead of transmission, it is clear that no information is transmitted.
- If, however, each bit is independently equally likely to be 0 or 1, 1000 shannons of information (more often called bits) have been transmitted. Between these two extremes, information can be quantified as follows.
- If $\mathbb{X}$ is the set of all messages $\{x_1, ..., x_n\}$ that *X* could be, and $p(x)$ is the probability of some $x \in X$, then the entropy, *H*, of *X* is defined:[

$$H(X) = \mathbb{E}_X[I(x)] = -\sum_{x \in \mathbb{X}} p(x) \log p(x).$$

- The special case of information entropy for a random variable with two outcomes is the binary entropy function, usually taken to the logarithmic base 2, thus having the shannon (Sh) as unit:

$$H_{\mathrm{b}}(p) = -p \log_2 p - (1-p) \log_2 (1-p).$$

Joint entropy

- The *joint entropy* of two discrete random variables *X* and *Y* is merely the entropy of their pairing: (*X*, *Y*). This implies that if *X* and *Y* are independent, then their joint entropy is the sum of their individual entropies.
- For example, if (*X*, *Y*) represents the position of a chess piece — *X* the row and *Y* the column, then the joint entropy of the row of the piece and the column of the piece will be the entropy of the position of the piece.
    - Despite similar notation, joint entropy should not be confused with *cross entropy*.

$$H(X,Y) = \mathbb{E}_{X,Y}[-\log p(x,y)] = -\sum_{x,y} p(x,y) \log p(x,y)$$

Conditional entropy (equivocation)

- The *conditional entropy* or *conditional uncertainty* of *X* given random variable *Y* (also called the *equivocation* of *X* about *Y*) is the average conditional entropy over *Y*:
- Because entropy can be conditioned on a random variable or on that random variable being a certain value, care should be taken not to confuse these two definitions of conditional entropy, the former of which is in more common use. A basic property of this

$$H(X|Y) = \mathbb{E}_Y[H(X|y)] = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y) = -\sum_{x,y} p(x,y) \log p(x|y).$$

- The *conditional entropy* or *conditional uncertainty* of *X* given random variable *Y* (also called the *equivocation* of *X* about *Y*) is the average conditional entropy over *Y*:

- Because entropy can be conditioned on a random variable or on that random variable being a certain value, care should be taken not to confuse these two definitions of conditional entropy, the former of which is in more common use. A basic property of this form of conditional entropy is that:

$$H(X|Y) = H(X,Y) - H(Y).$$

Mutual information (Transinformation)

- *Mutual information* measures the amount of information that can be obtained about one random variable by observing another. It is important in communication where it can be used to maximize the amount of information shared between sent and received signals. The mutual information of *X* relative to *Y* is given by:

$$I(X;Y) = \mathbb{E}_{X,Y}[SI(x,y)] = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

- where SI (*S*pecific mutual *I*nformation) is the pointwise mutual information.
- A basic property of the mutual information is that

$$I(X;Y) = H(X) - H(X|Y).$$

- That is, knowing *Y*, we can save an average of *I*(*X*; *Y*) bits in encoding *X* compared to not knowing *Y*.
- Mutual information is symmetric:

$$I(X;Y) = I(Y;X) = H(X) + H(Y) - H(X,Y).$$

Kullback–Leibler Divergence (Information Gain):

- The *Kullback–Leibler divergence* (or *information divergence*, *information gain*, or *relative entropy*) is a way of comparing two distributions: a "true" probability distribution *p(X)*, and an arbitrary probability distribution *q(X)*.
- If we compress data in a manner that assumes *q(X)* is the distribution underlying some data, when, in reality, *p(X)* is the correct distribution, the Kullback–Leibler divergence is the number of average additional bits per datum necessary for compression.
- It is thus defined

$$D_{\mathrm{KL}}(p(X)\|q(X)) = \sum_{x\in X} -p(x)\log q(x) \; - \sum_{x\in X} -p(x)\log p(x) = \sum_{x\in X} p(x)\log\frac{p(x)}{q(x)}.$$

Coding theory

- Coding theory is one of the most important and direct applications of information theory.
- It can be subdivided into source coding theory and channel coding theory.
- Using a statistical description for data, information theory quantifies the number of bits needed to describe the data, which is the information entropy of the source.
- Data compression (source coding): There are two formulations for the compression problem:
    - lossless data compression: the data must be reconstructed exactly;
    - lossy data compression: allocates bits needed to reconstruct the data, within a specified fidelity level measured by a distortion function. This subset of information theory is called *rate–distortion theory*.
- Error-correcting codes (channel coding): While data compression removes as much redundancy as possible, an error correcting code adds just the right kind of redundancy (i.e., error correction) needed to transmit the data efficiently and faithfully across a noisy channel.

Product Cryptosystems

- Data encryption scheme in which the ciphertext produced by encrypting a plaintext document is subjected to further encryption.
- By combining two or more simple transposition ciphers or substitution ciphers, a more secure encryption may result.
- A cryptosystem ***S=(P,K, C,e,d)*** with the sets ofplaintexts ***P***, keys ***K*** and cryptotexts***C*** andencryption (decryption) algorithms ***e*** (***d***) is called **endomorphic** if ***P=C***.
- If $S_1=(P,K_1, P,e^{(1)},d^{(1)})$ and $S_2=(P,K_2, P,e^{(2)},d^{(2)})$ are endomorphic cryptosystems,then the**product cryptosystem** is
- $S_1 \ddot{A} S_2=(P,K_1 \ddot{A} K_2, P,e,d)$,
- where encryption is performed by the procedure
- $e_{(k1,k2)}(w) = e_{k2}(e_{k1}(w))$
- and decryption by the procedure
- $d_{(k1,k2)}(c) = d_{k1}(d_{k2}(c))$

Cryptanalysis

- Cr**yptanalysis** is the study of analyzing information systems in order to study the hidden aspects of the systems.
- Cryptanalysis is used to breach cryptographic security systems and gain access to the contents of encrypted messages, even if the cryptographic key is unknown.
- In addition to mathematical analysis of cryptographic algorithms, cryptanalysis includes the study of side-channel attacks that do not target weaknesses in the cryptographic algorithms themselves, but instead exploit weaknesses in their implementation.

Methods

- *Ciphertext-only*: the cryptanalyst has access only to a collection of ciphertexts or codetexts.

- *Known-plaintext*: the attacker has a set of ciphertexts to which he knows the corresponding plaintext.
- *Chosen-plaintext* (*chosen-ciphertext*): the attacker can obtain the ciphertexts (plaintexts) corresponding to an arbitrary set of plaintexts (ciphertexts) of his own choosing.
- *Adaptive chosen-plaintext*: like a chosen-plaintext attack, except the attacker can choose subsequent plaintexts based on information learned from previous encryptions. Similarly *Adaptive chosen ciphertext attack*.
- *Related-key attack*: Like a chosen-plaintext attack, except the attacker can obtain ciphertexts encrypted under two different keys. The keys are unknown, but the relationship between them is known; for example, two keys that differ in the one bit.