Methods of Cryptanalysis
Lecture 2

Lecture Notes for a course given by Alex Biryukov
at the Weizmann Institute of Science, Israel.
Notes taken by Safro Ilya.

March 23, 2000

Lecture 2

History and Definitions of Classical Ciphers

2.1 A Brief History of Cryptography

Cryptography is one of the oldest fields of technical study we can find records of, going back at least 4,000 years[1]. It probably began in Egypt, where hieroglyphics were used to decorate the tombs of deceased rulers and kings. These hieroglyphics told the story of the life of the king and proclaimed the great acts of his life. In the year 1900 B.C. one inscription in the tomb of the nobleman Khnumhotep II used some unusual hieroglyphic symbols here and there in place of the more ordinary ones. They were purposefully cryptic, but not apparently intended to hide the text. It appears that the intention was to make the text seem more regal and important. As time went by, these writings became more and more complicated, and eventually the people lost interest in deciphering them. The practice soon died out, however cryptology was born.

"In its first 3000 years, it did not grow steadily. Cryptology arose independently in many places, and in most of them it died the deaths of its civilization. In other places, it survived, embedded in a literature, and from this the next generation could climb to higher levels."[2]

For almost 1600 years, the entire history of ancient Egypt was unknown because no one could read a picture language of the hieroglyphics. If it weren't for a "hard copy" called the Rosetta Stone, the language would probably still be lost, and we would know almost nothing of the first major civilization and world power on Earth. In 1799 soldiers in Napoleon's army were breaking down a wall in an old fort in Rosetta, Egypt. Suddenly, they came across a black slab of stone divided into three horizontal bands. Each band was covered with carved writing. The bottom band was in ancient Greek, the middle band in Demotic, an Egyptian "peoples" language, and the top band contained hieroglyphics, the language which literally covers the sacred temples and statues of ancient Egypt. Written in 196 BC, the stone was a record of the priestly honors and economic privileges given to Ptolemy V for the services he performed for Egypt.

Fortunately, ancient Greek is still well understood by scholars. When translators examined the last two lines in the Greek section, they found that the stone carried the same message in the three languages. By comparing the known Greek section with the characters in the hieroglyphic section, it became possible to decipher the hieroglyphic symbols.

In September 1822, Jean-Francois Champollion uncovered the techniques which led to our

[1]Most of the historic material in this section is taken from [2].

current understanding of hieroglyphics[2]. The Rosetta Stone is now located at the British Museum in London, England. The next lecture will briefly demonstrate the main idea of decipherment of the hieroglyphics.

The ancient Chinese used the ideographic nature of their language to hide the meaning of words. Messages were often transformed into ideographs for privacy, but no substantial use in early Chinese military conquests are apparent. Genghis Khan[3], for example, seems never to have used cryptography. In India, secret writing was apparently more advanced, and the government used secret codes to communicate with a network of spies spread throughout the country. Early Indian ciphers consisted mostly of simple alphabetic substitutions often based on phonetics. Some of these were spoken or used as sign language. This is somewhat similar to "pig latin" (igpay atinlay) where the first consonant is placed at the end of the word and followed by the sound "ay". The cryptographic history of Mesopotamia was similar to that of Egypt, in that special phonetically equivalent cuneiforms were used to complicate the text. This technique was also used in Babylon and Assyria. In the Bible, a Hebrew ciphering method is used at times. In this method, the last letter of the alphabet is replaced by the first, and vice versa, continuing to swap the letters in a mirror-like fashion. This is called 'atbash' and it was used by prophet Jeremiah (25:26, 51:41). Thus *Babel* (Hebrew word for "Babylon") is encrypted as *Sheshach*.

In the famous Greek drama the 'Iliad', secret writing was used when Bellerophon was sent to the king with a secret tablet which told the king to have him put to death. The king tried to kill him by having him fight several mythical creatures, but he won every battle. The Spartans used a system which consisted of a thin sheet of papyrus or leather wrapped around a staff (now called a "staff cipher"). Messages were written down the length of the staff, and the papyrus was unwrapped. In order to read the message, the papyrus had to be wrapped around a staff of equal diameter. Called the 'skytale' cipher, this was used in the 5th century B.C. to send secret messages between Greek warriors. Without the right staff, it would be difficult to decode the message using the techniques available at that time. The following version of the alphabet demonstrates the technique. First we see the wrapped version of the alphabet, then the unwrapped version.

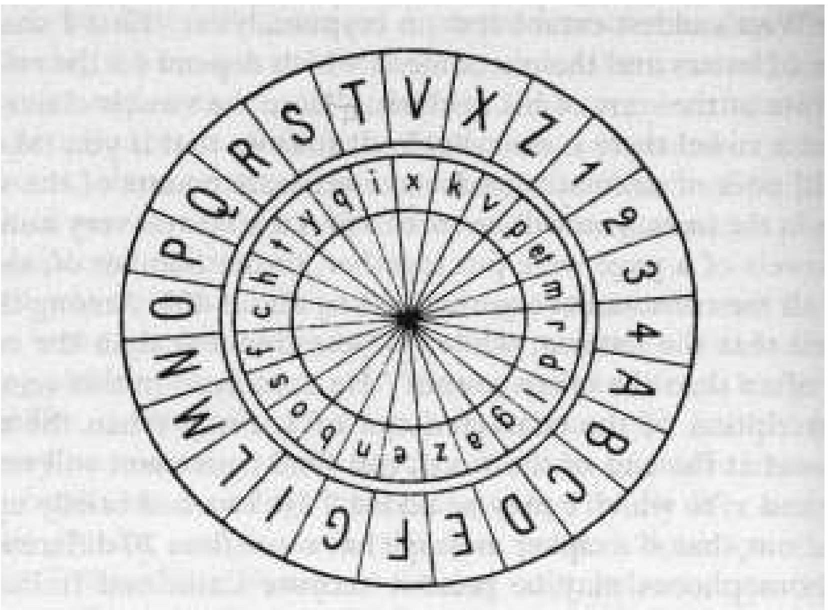
```
Plaintext   : ABCDEFGHIJKLMNOPQRSTUVWXYZ

     ADGJMPSVY
     BEHKNQTWZ
     CFILORUX

 Ciphertext : ADGJMPSVYBEHKNQTWZCFILORUX
```

Another Greek method was developed by Polybius (now called the "Polybius Square"). The letters of the alphabet would be laid out in a five by five square (similar to the later Playfair method) with i and j occupying the same square. Rows and columns are numbered 1 to 5 so that each letter has a corresponding (row,column) pair. These pairs could easily be signaled by torches or hand signals. Decryption consists of mapping the digit pairs back into their corresponding characters. This system was the first to reduce the size of the symbol set, and in a loose sense it might be considered the forerunner of modern binary representations of characters. Another digital method will appear later in this lecture.

[2]His work was based on an earlier breakthrough by Thomas Young (1814). Do you recall the modulus of Young, the wave theory of light, discovery of interference? This Young.

[3]The Mongol Conqueror of Asia, used many of the discoveries of Chinese wise men (powder, siege towers, etc.)

During the Middle Ages, cryptography started to progress. All of the Western European governments used cryptography in one form or another, and codes started to become more popular. Ciphers were commonly used to keep in touch with ambassadors. The first major advances in cryptography were made in Italy. Venice created an elaborate organization in 1452 with the sole purpose of dealing with cryptography. They had three cipher secretaries who solved and created ciphers that were used by the government. Leon Battista Alberti, who is best known as a Florentine architect, is also known as "The Father of Western Cryptology" in part because of his development of polyalphabetic substitution. He suggested to change the cipher alphabet every few words by rotating the *Alberti disk* to a new position (see Figure 2.1).



Leon Battista Alberti's cipher disk

Figure 2.1:

One of the most famous magic manuscripts, the so-called Leiden papyrus, discovered at Thebes and written in the third century A.D. in both Greek and very late form of demotic (a highly simplified version hieroglyphics) employs cipher to conceal the crucial portions of important recipes. The plaintext in most of the cipher sections is in Greek, and the cipher alphabet consists basically of Greek letter signs. Cryptology served magical purposes frequently throughout the Middle Ages, and even in the Renaissance was still disguising important parts of alchemical formulas. A manuscript compiled at Naples between 1473 and 1490 by *Arnaldus de Bruxella* uses 5 lines of cipher to conceal the crucial part of the operation of making a philosopher's stone.

The next major step was taken in 1518, by Trithemius, a German monk who had a deep interest in the occult. He wrote a series of six books called 'Polygraphia', and in the fifth book, devised a table that repeated the alphabet with each row a duplicate of the one above it, shifted over one letter (see Figure 2.6). To encode a message, the first letter of the plaintext is enciphered with the first row of the table, the second letter with the second row, and so on. This produces a message where all available ciphers are used before being repeated.

The most famous cryptographer of the 16th century was Blaise de Vigenere (1523-1596). In 1585, he wrote 'Tracte des Chiffres' in which he used a Trithemius table, but changed the way the key system worked. One of his techniques used the plaintext as it's own key. This method is called an "auto key" and is still used in modern ciphers. Another used the ciphertext.

From the Renaissance let us move to the end of 18th century. In the British colonies in North America, there was no centralized cryptographic organization. Decryption was done predominantly by interested individuals and men of the cloth. In 1775, a letter intercepted from Dr. Benjamin Church was suspected to be a coded message to the British, yet the American revolutionaries could not decipher it. Their problem was solved by Elbridge Gerry, who later became the fifth Vice-President, and Elisha Porter. The message proved Church guilty of trying to inform the Tories, and he was later exiled. Another spy, Benedict Arnold used a code wherein each correspondent has an exact copy of the same 'codebook'. Each word of plaintext is replaced by a number indicating its position in the book (e.g. 3.5.2, means page 3, line 5, word 2). An obvious drawback of this system was that one had to search through many pages in order to find a specific word. Due to this problem the codebook was later changed for a dictionary. Nevertheless, in-spite of intricate cryptography, Arnold's correspondent was caught and hung, (even before any messages were decrypted) so the codebook wasn't used much. The revolutionaries also employed ciphers during the war. Samuel Woodhull and Robert Townsend supplied General George Washington with much information about British troop strength and movements in and around New York City. The code they used consisted of about 800 numbers which replaced plaintext words. For example 711="George Washington", 356="letter", etc. This code was written by Major Benjamin Tallmadge. For further assurance, they also used invisible ink.

The use of cryptographic machines in 20-th century dramatically changed the nature of cryptography and cryptanalysis. Cryptography became intimately related to machine design, and security personnel became involved with the protection of these machines. The basic systems remained the same, but the method of encryption became reliable and electro-mechanical.

In 1929, Lester S. Hill published an article "Cryptography in an Algebraic Alphabet" in "The American Mathematical Monthly". Each plaintext letter was given a numerical value. He then used polynomial equations to encipher plaintext, with values over 25 reduced modulo 26. To simplify equations, Hill transformed them into matrices, which are more easily multiplied. This method eliminates almost all ciphertext repetitions, and is not broken with a normal frequency analysis attack. It has been found that if a cryptanalyst has two different ciphertexts from the same plaintext, and if they use different equations of the same type, the equations can be solved, and the system is thus broken. To counter charges that his system was too complicated for day to day use, Hill constructed a cipher machine for his system using a series of geared wheels connected together. One problem was that the machine could only handle a limited number of keys, and even with the machine, the system saw only limited use in the encipherment of government radio call signs. Hill's major contribution was the use of mathematics to design and analyze cryptosystems.

In the next section we will focus on some of the most popular and interesting simple ciphers, which were used through the long history of cryptography. Although weak by modern standards, when standing alone, some of these are important elements in modern cryptosystems.

2.2 Kerkhoff's design rules

Like any other theory, Cryptography has the number of basic rules and principles. *Auguste Kerkhoff*, a German philologists and mathematician was first, who designed the main principles for cryptosystems. In 1883 A. Kerkhoff published his work *"La Cryptographie Militarie"*, where he put forward the following six rules for design of useful cipher system :

1. The system should be unbreakable in practice, even it is theoretically breakable (this we would call now computational security vs. information theoretic security).
2. **The system should not be compromised if the opponent learns the technique being used. In other words, the security should reside in the choice of the key**.
3. The key should be easily remembered and changed.
4. The symbols of cryptogram should be convenient for transmition (for example by telegraph, or ASCII in modern computers).
5. Encryption apparatus is portable and allows use by a single operator.
6. The system should be easy to use (no long lists to remember, no mental strain, etc.).

As we see, these principles are rather philosofical than mathematical, but there are many cases in world history, when non-observance of these rules gave the serious consequences. For example in 1914-1917, the ciphersystem used by the Russian Tsar's army was so complex that under the stress of battle, the soldiers which were responsible for signal transmition threw it away and communicated in clear. This was the cost of the battle of Tannenberg.

2.3 Transposition ciphers

Transposition ciphers are block ciphers that change the positions of characters of message plaintext. The oldest known transposition cipher is the Scytale cipher used by ancient Greeks which was already mentioned before. One typical transposition cipher belonging to this category is the Railfence cipher used during the American Civil War.

```
T   C   E   S   D   I   W
 H S I H R A U E I C V L A
  I   P   W   S   N   I   R.

Ciphertext : TCESDIW HSIHRAUEICVLA IPWSNIR
```

This method is also known as depth-k cipher. Here plaintext pattern is rearranged in k-row decomposition to conceal the message. Here we use depth-3 row decomposition. Here the method of encoding is the key for a system.

A little bit more complicated was a method called **Redefence cipher**. Here the order of rows is defined using some numeric key, which numbers are according to the order of the rows in railfence cipher. With the previous plaintext and the key 312 we receive the following ciphertext :

```
Ciphertext : IPWSNIR TCESDIW HSIHRAUEICVLA
```

Next transposition method called a **grille cipher**. Grille is a square or rectangular piece of cardboard or other material into which openings have been cut at various places. The basic idea of using a grille for transposing letters or words goes back to *Girolamo Cardano*, a 16-th century Italian mathematician. The letters of a plaintext were written through the "windows", the grille was removed and the spaces between those letters filled with null letters to make a false message. When a similar grille was placed over the coded message, the null letters were hidden and the true letters could be read easily through the holes. This method was used later by Cardinal Richelieu.

Later, grilles were invented which could be rotated so that every cell of a matrix contained a letter of the original message, but in a mixed-up order. Such grilles were used by Germans in World War I [2, 3]. Each grille had a codename : ANNA for 25 letters, BERTA for 36, CLARA for 49, DORA for 64, EMIL for 81, FRANZ for 100. French cryptanalysts soon devised attacks exploiting weaknesses of the intriguing geometry, symmetries of the grilles. The grilles lasted four months.

5	6	7	8	9	5
9	2	3	4	2	6
8	4	1	1	3	7
7	3	1	1	4	8
6	2	4	3	2	9
5	9	8	7	6	5

HOW CELLS ARE NUMBERED FOR MAKING A GRILLE

Figure 2.2: Grille numbering.

To make a 6-by-6 grille, rule a square of cardboard into the 36-cell checkerboard shown in the first figure. Number the cells as indicated. Cut out any one cell with the number 1, another cell with 2, and so on until the square has nine cells missing. For instance, the grille may look like on shown in the second figure.

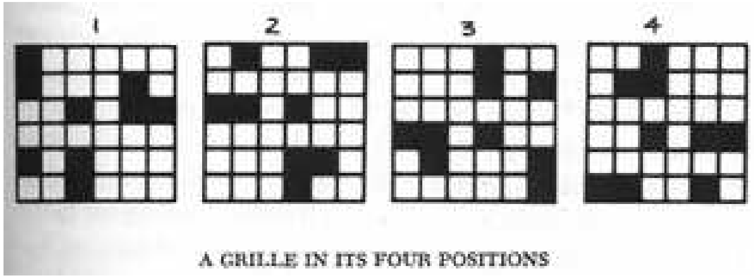


Figure 2.3: Grille cipher.

We will call missing cell a hole or window even though two or more open cells may be next to each other, forming a single opening. Draw a square of the same size on a sheet of blank paper. Place the grille on this square in any position and write the first nine letters of the message in the nine holes. Now give the grille a quarter-turn clockwise and write the next nine letters. Another quarter-turn in the same direction permits to add nine more letters, and a final turn allows nine more. Now remove a grille and we got a matrix with ciphertext, which could be send to recipient in this or any other form. It is easy to check, that if the enemy knows the method of encryption and uses exhaustive search, it would take about 2^{18} steps for 6×6 grille and if we will take 10×10 grille, the exhaustive search decryption would take about 2^{50} steps.

Another way to encrypt the message using transposition methodology is a **Nihilist transposition cipher** which was invented by Russian nihilists [4]. Here we need a $n \times n$ square. The same key of length n is applied to rows and columns. We write the plaintext inside this square and each letter is entered into a second square, which contains the ciphertext, according to its numerical coordinates.

```
Plaintext : nihilist transposition
Key : 23145

  2 3 1 4 5            1 2 3 4 5
2 n i h i l          1 s a n p o
3 i s t t r  ===>>>  2 h n i i l
1 a n s p o          3 t i s t r
4 s i t i o          4 t s i i o
5 n . . . .          5 . n . . .

Ciphertext : sanpohniiltistrtsiio.n...
```

As we see in general we can define any permutation on the string of plaintext and receive the transposition encryption.

2.4 Substitution ciphers

Encryption using *Substitution* cipher is a one of the most simple methods used even by ancient cryptographers. The main idea of Substitution cipher system is substitution of one character in the plaintext by another one in the ciphertext. Before we explain, let us introduce a notation: we will denote a cipher's alphabet as Σ and the size of the alphabet as $|\Sigma|$. There are four basic types of substitution ciphers (see Figure 2.4 for a classification of substitutions):

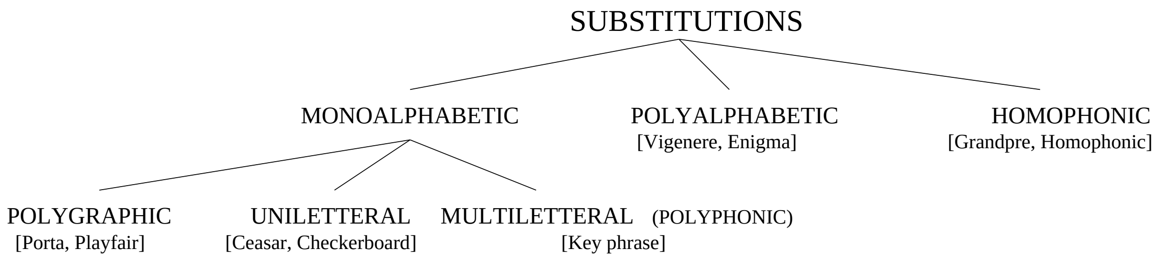


Figure 2.4: Classification of substitution ciphers, with examples for each class.

1. **Monoalphabetic substitution**. The main idea of these ciphers is a one by one substitution of a symbol from plaintext to corresponding symbol in ciphertext. Usually this method is used to encrypt the entire plaintext. If we will take an English alphabet as alphabet for plaintext then we will get $26! \approx 4 \cdot 10^{26}$ possible different ciphers corresponding to permutations in English. For example the cipher which uses shifting and multiplication is forming an affine

[4] An anarchist trend in Russian society in XIX century

transformation given two keys :

$$E(x) = ax + b \ (mod|\Sigma|)$$

where a and b are the keys. Another way to represent the general case of monoalphabetic cipher using n keys is :

$$E(x) = poly(x) \ (mod|\Sigma|)$$

In general simple substitution may have no structure and be an arbitrary permutation of letters. Any such a method of encryption could be attacked by using a frequency table of letters of the corresponding alphabet.

2. **Homophonic substitution**. This substitution maps one symbol from the plaintext alphabet to the subset of ciphertext's alphabet symbols using some rule. This one-to-many mapping is harder to break than monoalphabetic version of substitution cipher. Homophonic substitution ciphers were used as early as 1401 by the Duchy of Mantua [4].

3. **Polyalphabetic substitution**. This is a method to use multiple monoalphabetic ciphers. For example, if we decide to use n monoalphabetic ciphers to encode some plaintext, in every moment we must decide what cipher should be used for now. Such a decision might be dependent of position of currently encrypted symbol in plaintext. *Leon Alberti* (1568) was the first who invented this method. Such ciphers became very popular due to their superiority over the simple substitution methods. For example, one of them was used by the Union army during the American Civil War (confederates used the railfence cipher, described above).

   Most polyalphabetic ciphers are periodic substitution ciphers based on a period p. For instance, if a plaintext is

   $$X = x_1x_2...x_px_{p+1}...x_{2p}...$$

   then the ciphertext after encryption should be

   $$E_k(X) = e_1(x_1)e_2(x_2)...e_p(x_p)e_1(x_{p+1})...e_p(x_{2p})...$$

   where e_i is a monoalphabetic cipher.

4. **Polygram cipher**. In the polygram cipher, several characters are encrypted at the same time using one key. These ciphers take the blocks of symbols and encrypt them to blocks of other symbols. For example "MAM" could correspond to "ZXY".

Let us provide some examples of monoalphabetic substitution ciphers.

- **Caesar cipher**. As was mentioned in Kerchoff's principles, "the key should be easily remembered". Probably the understanding of this rule lead Caesar to invent his well known cipher. Denote by Σ an alphabet for plaintext and by Ω an alphabet of ciphertext. Caesar cipher is a fixed-key monoalphabetic substitution cipher which takes as a key an integer number $k \leq |\Sigma|$ and substitutes i-th letter in the plaintext by $(i+k)$-th letter modulo $|\Sigma|$. Julius Caesar took $k = 3$.

```
Plaintext  : PLAINTEXT
Key        : k=3 and English alphabet
Ciphertext : SODLQWHAW
```

Notice, that there are only 26 possible keys in this cipher.

- **Baconian cipher**. Francis Bacon (1561-1626) and his Rosicrucian fraternity made use of several different kinds and types of ciphers, some of them to sign various published works issued under different names or pseudonyms, and some of them to give messages or teachings. Here we introduce one of his simplest ciphers in which the cipher equivalents are five-letter groups and the resulting cipher is monoalphabetic in character. Bacon uses a 24-letter cipher with I and J, U and W used interchangeably.

```
A =  aaaaa      I/J  = abaaa       R   = baaaa
B =  aaaab       K   = abaab       S   = baaab
C =  aaaba       L   = ababa       T   = baaba
D =  aaabb       M   = ababb      U/V  = baabb
E =  aabaa       N   = abbaa       W   = babaa
F =  aabab       O   = abbab       X   = babab
G =  aabba       P   = abbba       Y   = babba
H =  aabbb       Q   = abbbb       Z   = babbb
```

Replace each letter with its Baconian equivalent.

```
Plaintext:   s     u     c     c     e     s     s
           baaab baabb aaaba aaaba aabaa baaab baaab
```

So far this is just a method of a binary encoding for the letters (substitute a=0, b=1). The main novelty was since the a-units and b-units are concealed; in this example the initial letter of each word indicates a or b: A-M = a, N-Z = b. To construct the ciphertext we must write the sentence which first letters of all words converted to a, b according to our rule give the word "success" after decryption.

```
               S = baaab

               b  a   a       a   b
Ciphertext:    We are looking for you.
```

This way a secret message can be hidden in any passage of the text. There is a popular enigmatologist trend to searching for Baconian ciphers in Shakespear's works in order to prove autoship of Francis Bacon for them.

- **Polybius square**. Polybius was an ancient Greek writer who proposed a method of substituting different two-digit numbers for each letter. The alphabet is written inside a 5-5 square matrix which has numbered rows and columns :

```
   1   2   3   4   5
   -----------------
1| A   B   C   D   E
2| F   G   H  I/J  K
3| L   M   N   O   P
```

```
4| Q   R   S   T   U
5| V   W   X   Y   Z
```

To encode, substitute for each letter the numbers marking the coordinates of the letter in this matrix. Always put the row number first. For example "F" would be equal to 21 and the word "RENAISSANCE" would be encoded as "42153311244343113313 15". It is more a code than a cipher, and it was used to transmit "long-distance" messages at night using torches.

- **Checker board**. This method of encryption is similar to Polybius square. Here the key contains three words . Two of them placed on boundary of a 5-5 matrix and form the coordinates of the entries. These words can not contain similar letters in order to provide unique coordinates in the matrix. Inside of the matrix we write the alphabet. However we would like to fool the enemy and to achieve this we use the third keyword. We write it into the matrix from its first cell and fill the other cells with remaining letters of alphabet. The encryption algorithm is the same as in Polybius square.

```
    G R O U P
   ----------
W|  B O A R D
H|  E F G H I
I|  K L M N P
T|  Q S T U V
E|  W X Y Z C

 Plaintext  : protocol
 Ciphertext : IPWUWRTOWREPWOIR
```

The substitution ciphers in the previous examples were monoalphabetic, that means one symbol for each letter. Such ciphers are not difficult to solve, even if the plaintext is relatively short (70–100 symbols). Next level of difficulty of substitution ciphers is called digraphic or polygraphic ciphers. In these cipher symbols of plaintext alphabet are encrypted in pairs (digraphs) thus making a frequency analysis much harder. Let us focus on a few examples:

- **Porta's (1563) digraphic cipher**. In the Porta Cipher, a single symbol is substituted for every pair of letters in the plaintext. It is a first known digraphic cipher. To use this cryptosystem we need a 26×26 matrix, the entries of which should be filled with different symbols (can be just numbers). The alphabet is written outside the border, once from top to bottom and once from left to right of the matrix. Instead of strange-looking characters of the original Porta's table (see Figure 2.5) one can use numbers 1...676, placed at random in the table. Suppose we want to encode the word "LETTER". We take the first pair of letters in our word "LE", look at "L" as a y-coordinate and "E" as a x-coordinate in this matrix and search for the entry with coordinates ("L", "E"). Let the entry be 291. Similarly to "LE", we will look for "TT" and "ER". Now concatenate the resulting numbers and you will get the ciphertext. In this case it would "291-514-122". Notice that the original Porta's table possesses some structure. The intention was to help decipher the cryptograms, however such structure also offers considerable help to the attacker. One possible solution to this problem

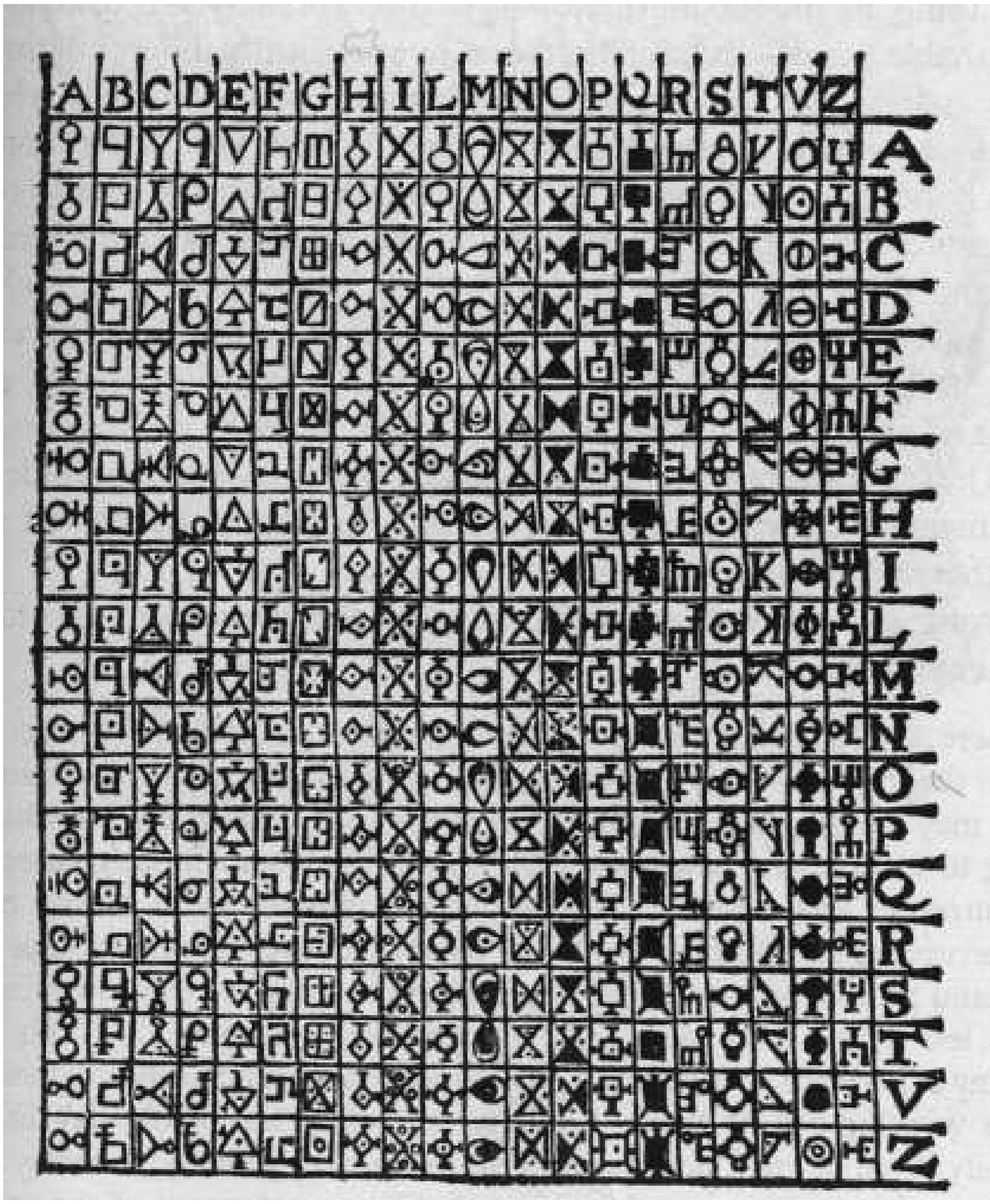


Figure 2.5: Porta's digraphic table (1563).

is to leave an ordered set of symbols (for example numbers) in the table, but to reorder the alphabet's that index the table.

- **The Playfair cipher** This is a more compact digraphic cipher than Porta's cipher. The system contains a 5×5 matrix. In the first row we write into this matrix a keyword, which contains only different letters. The remaining cells we fill with the rest of letters in some easy to remember order. As in Porta's cipher, the message is encoded by taking pairs of letters. There are three basic rules :

  1. If both letters happen to be in the same row, use the letters immediately to the right of each letter. Think of the right end of each row as being joined to its left end.
  2. If both letters are in the same column, use the letters immediately below each letter. Think of the bottom of each column as connected to its top.
  3. If the two letters are in different rows and in different columns, each letter is replaced by the letter in the same row that is also in the column occupied by the other letter.
  4. In case of two equal letters in a digram, insert 'x' between them (thus "choose" will become "choxose").

```
  1  2  3  4  5
1 P I/J A  N  O
2 B  C  D  E  F
3 G  H  K  L  M
4 Q  R  S  T  U
5 V  W  X  Y  Z

Plaintext  : cipher
Key        : piano
Encryption : "cr" -> "hc"
             "ph" -> "ig"
             "er" -> "ct"
Ciphertext : "hcigct"
```

As we can see from these first simple examples, the advantages of digraphic ciphers are clear : The frequency distribution becomes much more flat, variance becomes higher since the alphabet of the cipher is now 676 letters. The second advantage is that the length of ciphertext is twice shorter than in plaintext.

Next large class are the polyalphabetic ciphers :

- **Vigenere and Beaufort ciphers**. A popular form of polyalphabetic cipher based on shifted alphabets is the Vigenere cipher invented by French cryptologist *Blaise de Vigenere* (1523). The Vigenere cipher is based on the Vigenere (Trithemius) tableau (See Figure 2.6). Here we start by describing a simpler (and weaker) variant, which was suggested by Lewis Caroll. To encode using this cipher, let us choose a keyword "VIGENERE", split a plaintext into strings of length equal to the length of the keyword and place the bellow the keyword in one column of the same length. The encoding should be as following : all letters in the first column are enciphered using "V" as the key, in the second using "I", etc. Thus to encipher the first letter

"T" of plaintext, look down the K (key) column of the tableau (bellow) for "V" and across the top row for "T". Where "V"'s row meets "T"'s column find the substitute "O".

So, denote a plaintext message as $M = m_1 m_2 ... m_n$ and let us define the key K by a sequence $K = k_1 ... k_d$, where $k_{i \ mod \ |\Sigma|}$ represents a shift for i-th alphabet :

$$c_i \ = \ e_i(m_i) \ = \ (m_i \ + \ k_{i \ mod \ |\Sigma|}) \ mod \ |\Sigma| \ .$$

And the decryption would be :

$$m_i \ = \ d_i(c_i) \ = \ (k_{i \ mod \ |\Sigma|} \ + \ c_i) \ mod \ |\Sigma| \ .$$

The **Beaufort cipher** is a variant of the above scheme :

$$e_i(m_i) \ = \ (k_{i \ mod \ |\Sigma|} \ - \ m_i) \ mod \ |\Sigma| \ .$$

And the same function can be used to decipher; that is for ciphertext letter c :

$$d_i(c) \ = \ (k_{i \ mod \ |\Sigma|} \ - \ c) \ mod \ |\Sigma| \ .$$

So the Beaufort cipher reverses the letters in the alphabet, and then shifts them to the right by k_i positions. A nice feature of Beaufort is that encryption and decryption are performed by the some procedure. The cipher is so simple that it was sold together with postcards in England in order to allow discrete communications.

- **Autokey**. This is a method of generating the key during the coding process. The keyword string starts from a real keyword and concatenate it with the plaintext. Using a previous Vigenere tableau :

```
Key    : p i a n o t o b e o r n o t t o b e
Plain  : t o b e o r n o ...
Cipher : i w b r c k b p ...
```

  The main observation that can be done here is that the key is the function of plaintext.

Now let us consider another class of substitution ciphers: **homophones**. As we said a homophonic substitution cipher maps each character a of the plaintext alphabet in a set of ciphertext elements $f(a)$ called homophones. A plaintext message $M = m_1 m_2 ...$ is encrypted as $C = c_1 c_2 ...$, where each c_i is picked from the set of homophones $f(m_i)$, at random or by some procedure.

Homophonic ciphers are much more difficult to solve than simple substitution ciphers, especially when the number of homophones of a letter is proportional to the relative frequency of the letter. This is because the relative distribution of the ciphertext symbols will be nearly flat, confusing frequency analysis.

- **Grandpre cipher**. This is a typical homophonic cipher, which extends the alphabet of plaintext and using some rule one letter from Σ should have more than one homophone. We define 8×8 matrix and fill this matrix with keywords (might be a sentence). Let "T" be the first letter of plaintext. Using simple lexicographic order of coordinates, we search for a first occurrence of "T" and assign 26 to be the first symbol of ciphertext and so on until we meet in plaintext the letter that was already occurred. In this case this is "T". We search the second letter "T" in ordered list of coordinates and assign now $T \ = \ 84$.

```
Vigenere tableau :

K (key) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z  (plaintext)
------- ---------------------------------------------------       P
  A     A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
  B     B C D E F G H I J K L M N O P Q R S T U V W X Y Z A
  C     C D E F G H I J K L M N O P Q R S T U V W X Y Z A B
  D     D E F G H I J K L M N O P Q R S T U V W X Y Z A B C
  E     E F G H I J K L M N O P Q R S T U V W X Y Z A B C D
  F     F G H I J K L M N O P Q R S T U V W X Y Z A B C D E
  G     G H I J K L M N O P Q R S T U V W X Y Z A B C D E F
  H     H I J K L M N O P Q R S T U V W X Y Z A B C D E F G
  I     I J K L M N O P Q R S T U V W X Y Z A B C D E F G H
  J     J K L M N O P Q R S T U V W X Y Z A B C D E F G H I
  K     K L M N O P Q R S T U V W X Y Z A B C D E F G H I J
  L     L M N O P Q R S T U V W X Y Z A B C D E F G H I J K
  M     M N O P Q R S T U V W X Y Z A B C D E F G H I J K L  ciphertext
  N     N O P Q R S T U V W X Y Z A B C D E F G H I J K L M      C
  O     O P Q R S T U V W X Y Z A B C D E F G H I J K L M N
  P     P Q R S T U V W X Y Z A B C D E F G H I J K L M N O
  Q     Q R S T U V W X Y Z A B C D E F G H I J K L M N O P
  R     R S T U V W X Y Z A B C D E F G H I J K L M N O P Q
  S     S T U V W X Y Z A B C D E F G H I J K L M N O P Q R
  T     T U V W X Y Z A B C D E F G H I J K L M N O P Q R S
  U     U V W X Y Z A B C D E F G H I J K L M N O P Q R S T
  V     V W X Y Z A B C D E F G H I J K L M N O P Q R S T U
  W     W X Y Z A B C D E F G H I J K L M N O P Q R S T U V
  X     X Y Z A B C D E F G H I J K L M N O P Q R S T U V W
  Y     Y Z A B C D E F G H I J K L M N O P Q R S T U V W X
  Z     Z A B C D E F G H I J K L M N O P Q R S T U V W X Y


  Key        : VIGENERE
               --------
  Plaintext  : TOBEORNO
               TTOBE
               --------
  Ciphertext : OWHIBVES
               OBUFR
```

Figure 2.6: Example of Vigenere encryption.

```
   1 2 3 4 5 6 7 8
   ---------------
1| a b a s h i n g
2| y o k o h a m a
3| c o e x i s t s
4| d e a t h f u l
5| j a c k p o t s
6| q u i v e r e d
7| w i t c h i n g
8| z o d i a c a l

Plaintext  :  T  H  I  S  I  S  A  P  L  A  I  N  T  E  X  T
Ciphertext : 44 15 16 14 35 38 11 55 48 43 72 77 73 33 34 44
```

- **Beale cipher**. This cipher was invented by *T. J. Beale* (1820). The key for this cipher is the Declaration of Independence. The words of the key are numbered consecutively as shown bellow.

```
(01) When, in the course of human events, it becomes necessary
(11) for one people to dissolve the political bands which have
(21) connected them with another ...
                                .....
(99) these are Life, Liberty, and the pursuit of Happiness.
```

  Beale enciphered each letter in the plaintext message by substituting the number of some word which started with that letter. The letter "W", for example, was enciphered with the numbers 1, 19, 40, 66, 72, 290, 459.

Polyphonic (or multi-letteral) substitution is a kind of substitution in which a number of different letters of plaintext alphabet might be assigned to the same letter in alphabet of ciphertext.

- **Keyphrase cipher**. The keyword for this cipher is a 26-letter phrase which place below the alphabet and encoding is performed with the assignment of each letter of the plaintext alphabet with corresponding letter from this keyword phrase.

```
P:  a b c d e f g h i j k l m n o p q r s t u v w x y z
K:  n u m b e r s r e p r e s e n t i n g m e a n i n g

P:  h o w a r e y o u
C:  r n n n n e n n e
```

Bibliography

[1] Dorothy Elizabeth Robling Denning, *Cryptography and Data Security*, Addison-Wesley Publishing Company, Inc. 1983.

[2] D. Kahn, *The Codebreakers*, Macmillan, 1967.

[3] Martin Gardner, *Codes, Ciphers and Secret writing*, Dover Publications Inc. 1972.

[4] Man Young Rhee, *Cryptography and Secure Communications*, McGraw-Hill Series on Computer Communications 1994.

[5] Bruce Schneier, *Applied Cryptography*, John Wiley and Sons Inc. 1994.