CODES AND CIPHERS (Cryptology)

For deposit in the NSA Library — that great repository of cryptologic lore.
William F. Friedman
Washington
1956

by

WILLIAM F. FRIEDMAN

This is a revision of my original article, prepared in 1927 for the 14th Edition.
W.F.F.

from Encyclopaedia Britannica

CODES AND CIPHERS (CRYPTOLOGY), general terms designating the methods or the paraphernalia employed in secret communications or involved in the science of cryptology (from Gr. *kryptos,* "hidden," and *logos,* "word"). Because of the growth of governments, the expansion of commerce and especially the remarkable progress made in communications-electronics technology, cryptology has come to play a very important role in governmental communications, especially diplomatic and military. It also plays a minor role in commercial, industrial and banking communications. Among the more uncommon uses of cryptology are those in connection with attempts to establish authorship in cases where that has been brought into question, as, for example, that of the Shakespeare plays.

In its early stages cryptology was concerned almost exclusively with secrecy in written communications and this article will be restricted very largely thereto, but the science has developed to the stage where it deals not only with enciphered writing (cryptograms) but also with other mediums of cryptocommunication, such as enciphered telephony (ciphony) and enciphered facsimile (cifax) transmissions.

Cryptology embraces the twin or complementary sciences of signal security and signal intelligence. The former deals with all the means and methods of protecting one's own signals against interception and reading or utilization by unauthorized persons generally referred to as "the enemy." The latter deals with all the means and methods employed in acquiring information or intelligence by intercepting and solving the enemy's cryptosignals or nullifying his signal security so that the signals or information derived from them can be used against him.

Signal Security.—The principal components of this phase of cryptology are: (1) physical and personnel security; (2) transmission security; and (3) cryptosecurity. The first deals with the precautions and measures taken to assure that the physical arrangements and the facilities or procedures for safeguarding the paraphernalia or cryptomaterials used, *i.e.,* the codes, ciphers, key lists, etc., are adequate for the purpose and that the personnel employed in operating the codes and ciphers or cipher machines are trustworthy. The second component deals with the means, methods and procedures for assuring that no information is inadvertently disclosed either by indiscretions of operators or by faults in the transmitting or receiving apparatus which may assist in the solution of the transmissions. The third component, cryptosecurity, which deals with the technical adequacy of the cryptosystems employed, is usually of greater interest and deserves more extensive treatment than the other two. In an article of this nature it is possible only to deal briefly with cryptography (from Gr. *kryptos,* "hidden," and *graphein,* "to write"), it being understood that many of the cryptoprinciples employed for the protection of written communications, or signals representing them, can also be used in protecting or disguising other types of cryptosignalling; *e.g.,* ciphony and cifax. Cryptography deals with the processes, methods or means involved in preparing cryptograms, that is, messages or writings which are intended to be incomprehensible except to those who legitimately possess the proper special paraphernalia and the keys for those cryptograms and know how to use them in order to reproduce the original plain text of the messages. These processes are usually accomplished by means of cryptosystems employing codes or ciphers. The process of converting a plain-text message into a cryptogram is called enciphering (or encoding); that of reconverting the cryptogram back into its intelligible form, when done by a legitimate or authorized communicator, *i.e.,* one who legitimately holds the paraphernalia and the key, is called deciphering (or decoding).

Although in theory no sharp line of demarcation can be drawn between code systems and cipher systems, in modern practice the technical differences between them are sufficiently marked to warrant their being treated as separate categories of methods. Some authors include as a third and separate category the extensive but much less important one containing the so-called "concealment systems," which are sometimes employed to hide an internal or secret message within an external or apparently innocent piece of writing with a view to avoiding arousing suspicion in the minds of persons not privy to the secret, or to eluding censorship in wartime. In such systems the message or its elements are hidden or disguised by any one of hundreds of different means and methods, including such mediums as secret or invisible inks, microscopic writing, etc., but none of these concealment systems or devices will even be mentioned again herein. It is convenient to consider cipher systems first, then code systems, with the understanding that only a very few of the limited number of systems suitable for serious usage can here be outlined.

Cipher Systems.—In general, cipher systems involve a cryptographic treatment of textual units of constant and equal length, usually single letters, sometimes pairs, rarely sets of three letters, these textual units being treated as symbols without reference to their identities as component parts of words, phrases and sentences. Every practical cipher system must combine (1) a set of rules, processes or steps constituting the basic cryptographic method of treatment or procedure, called the general system, which is agreed upon in advance by the communicators and which is constant in character, with (2) a specific key which is variable in character. In enciphering plain text, the specific key, which may consist of a number or a series of numbers or a word, phrase, sentence, etc., controls the steps under the general system and determines the specific nature or exact composition of the cipher.

Printed in U. S. A.

message produced; in decipherment the specific key similarly controls the steps and determines what the deciphered text will be. When all operations are performed correctly, the two plain texts (before and after the cryptography) should be identical or nearly so, save for minor differences arising from errors in their encipherment and transmission or in their reception and decipherment. The general system should be such that even if it is known to the enemy no properly enciphered message can be read by him unless he also knows the specific key or keys applicable to that message.

Despite a great diversity in the external appearance and internal constitution of ciphers, there are only two basic classes of systems—transposition and substitution. (Concealment ciphers are excluded from the discussion.) A transposition cipher involves a rearrangement or change in the sequence of the letters of the plain-text message without any change in their identity; a substitution cipher involves a replacement of the plain-text letters by other letters (or by other symbols) without any change in their sequence. The two systems may be combined in a single cryptosystem.

The majority of transposition systems involve inscribing the letters of the plain text in a geometrical design called a matrix, beginning at a prearranged initial point and following a prescribed route, and then transcribing the letters from the matrix, beginning at another prearranged initial point and following another prescribed route. The matrix may take the form of a rectangle, trapezoid, octagon, triangle, etc., but systems in which the specific keys consist solely in keeping the matrices, the initial points and the routes secret are not often now employed because of their limited variability and, therefore, their relatively low degree of security. In this same class also fall systems which employ perforated cardboard matrices called grilles, descriptions of which will be found in most of the older books on cryptography. The transposition system most commonly used in practice is that designated as columnar transposition, wherein the transposition matrix takes the form of a simple rectangular figure the dimensions of which are determined in each instance jointly by the length of the individual message and the length of the specific key. An example is shown in fig. 1.

Plain-text message: **DELAY DEPARTURE UNTIL FURTHER NOTICE.**

T	E	L	E	G	R	A	P	H
9	2	6	3	4	8	1	7	5
D	E	L	A	Y	D	E	P	A
R	T	U	R	E	U	N	T	I
L	F	U	R	T	H	E	R	N
O	T	I	C	E				

Key

Cryptogram

ENEET FTARR CYETE AINLU UIPTR DUHDR LO

Fig. 1.—An example of simple columnar transposition. The numerical key is derived by numbering the letters of the keyword ("telegraph") in accordance with their relative order of appearance in the ordinary alphabet, repeated letters, if present in the key, being numbered in sequence from left to right. The message is written in the normal manner from left to right in successive horizontal lines underneath the key, forming a "rectangle" of columns of letters. These letters are then transcribed in regular groups of five from the rectangle by reading down the columns, taking the latter in the sequence indicated by the key numbers. The last line of the rectangle may be completely filled with letters, nonsignificants being added if necessary; but the security of the method is considerably increased if the last line shows one or more blank spaces, as in this example.

In the foregoing case the letters undergo a single transposition; in cases involving double transposition, that is, wherein the letters undergo two successive transpositions, the security of the cryptograms is very greatly increased, provided the methods selected are such as will effectively disarrange individual letters and not merely whole columns or rows. A practical system of double transposition is illustrated in fig. 2. The principal advantages of transposition systems lie in their comparative simplicity, speed of operation and, in some cases, their high degree of security; but despite these important considerations they do not at the present time play a prominent role in practical cryptography.

Substitution systems involve the use of conventional or cipher alphabets composed of two juxtaposed sequences, one (either expressed or implied) corresponding to the letters of the ordinary alphabet (*a, b, c . . . x, y, z*), the other containing their respective

Plain-text message: **DELAY DEPARTURE UNTIL FURTHER NOTICE.**

T	E	L	E	G	R	A	P	H
9	2	6	3	4	8	1	7	5
D	E	L	A	Y	D	E	P	A
R	T	U	R	E	U	N	T	I
L	F	U	R	T	H	E	R	N
O	T	I	C	E				

(a)

T	E	L	E	G	R	A	P	H
9	2	6	3	4	8	1	7	5
E	N	E	E	T	F	T	A	R
R	C	Y	E	T	E	A	I	N
L	U	U	I	P	T	R	D	U
H	D	R	L	O				

(b)

Cryptogram

TARNC UDEEI LTTPO RNUEY URAID FETER LH

Fig. 2.—An example of true double transposition. The cipher letters resulting from the first transposition rectangle (a) are written under the key of the second transposition rectangle (b) just as though they constituted plain-text, that is, from left to right, in successive horizontal rows. The final transposition is then performed in exactly the same manner as in fig. 1, yielding five-letter groups. The two rectangles may be based upon the same keyword, as in this example, or upon wholly different keywords.

cipher equivalents. The complexity of a substitution system usually depends upon three factors: (1) the specific composition of the cipher alphabet or alphabets employed; (2) the number of them involved in a single cryptogram; and (3) the specific manner in which they are used. As to their composition, cipher alphabets are of various types and are known under various names, such as standard, direct, reversed, systematically mixed, key-word mixed, random mixed, reciprocal, etc., all having reference to the nature of the two sequences composing them, the interrelations existing among them internally or externally, etc. The most important factor in connection with a cipher alphabet is whether its two sequences, regardless of their composition, are known or unknown to the enemy; for, if known, any conventional or disarranged alphabet may be handled with the same facility as the normal alphabet. As to the number of alphabets involved in it, a cryptogram is either monoalphabetic, involving a single cipher alphabet, or polyalphabetic, involving two or more alphabets. In essence the difference between the two types lies in the fact that in the former the equivalence between plain-text and cipher letters is invariant, *i.e.*, the equivalence is of a constant or invariable nature throughout the cryptogram, whereas in the latter it is of a changing or variable nature, controlled by the key. With regard to secrecy, the third condition mentioned above, namely, the specific manner in which the various cipher alphabets are employed, is the most important in determining the degree of security or resistance the cryptogram will have against cryptanalysis, as explained below.

Monoalphabetic substitution is usually uniliteral; *i.e.*, each letter of the plain text is replaced by a single character. Cases of biliteral, triliteral, etc., substitution are sometimes encountered; an example using biliteral equivalents is shown in fig. 3. No matter how many characters are in the groups composing the cipher equivalents for plain-text letters, if the groups are always the same for each letter, the substitution system is still monoalphabetic and the cipher can be treated as such.

Polyalphabetic systems are often referred to as double-key systems and, as noted above, employ two or more cipher alphabets in the encipherment of single dispatches. In a given system the

Plain-text message: **DELAY DEPARTURE UNTIL FURTHER NOTICE.**

Second letter / First letter

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

Enciphering

Plain: D E L A Y D E P A R T
Cipher: CA AB AC BA ED CA AB BB BA AE AA

Plain: U R E U N T I L F U R
Cipher: DE AE AB DE DA AA CC AC CB DE AE

Plain: T H E R N O T I C E
Cipher: AA BC AB AE DA DB AA CC BE AB

Cryptogram

CAABA CBAED CAABB BBAAE etc.

Fig. 3.—An example of biliteral, monoalphabetic substitution. A keyword alphabet of 25 letters (I serving also for J) is written in a square 5 x 5. (In this case the alphabet is based upon the word "telegraph.)" The letters (A,B,C,D, E) at the side and top of the square, taken in pairs, can then be used to represent the letters within the square. Thus, D = CA, E = AB, etc. The letters at the side of the square may be the same as or different from those at the top; in both cases keywords, identical or different, may be used instead of the letters A,B,C,D,E.

cipher alphabets may be entirely independent or they may be interrelated. In the simplest case of interrelated alphabets, that resulting from juxtaposing, at all points of coincidence, two primary sequences which are normal sequences proceeding in the same direction ($a, b, c \ldots x, y, z$), the secondary cipher alphabets are called standard alphabets; these when successively tabulated yield a symmetrical or square table (fig. 4) which is generally referred to in the literature as a Vigenère table, a quadricular table, etc.

Such a table may be used in several ways. The most common method employs the top row of letters as the plain-text sequence and the successive horizontal rows as the cipher sequences, each of which may be designated by its initial letter. Thus, the fourth or D alphabet is the one in which plain text A is represented by cipher D; *i.e.*, $A_p = D_c$, $B_p = E_c$, etc. In the ninth or I alphabet, $A_p = I_c$, $B_p = J_c$, etc. Therefore, when several such alphabets are to be employed, the letters of a key word can serve to indicate the number, identity and sequence of the cipher alphabets to be used consecutively in the cryptographic operations.

This type of cryptogram is usually referred to as the Vigenère cipher or as *le chiffre indéchiffrable*, although Blaise de Vigenère never claimed that he invented the table or the system; he was merely among the first to describe it (*see* below).

Plain-text letter

Key letter	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
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
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
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

Fig. 4.—The simple Vigenère table. The cipher equivalent of a given plain-text letter enciphered by a given key letter is that letter which stands at the intersection of the vertical column headed by the former and the horizontal row begun by the latter. Thus, plain-text letter E enciphered by key letter M yields cipher letter Q, or, in brief notation, $E_p(M_k) = Q_c$; $E_p(X_k) = B_c$, etc.

Various modifications of the simple square table are encountered, these involving different types and arrangements of the sequences therein, but in every case where these sequences can be arranged so as to exhibit symmetry of position as regards the order in which the letters fall in the successive rows or columns, sliding or concentric primary sequences may be employed to produce the same results as the table and are often more convenient.

Polyalphabetic substitution systems may be classified as periodic or aperiodic, depending upon whether or not the cryptograms produced by them exhibit external phenomena of a cyclic nature. Periodicity is exhibited externally whenever the substitution process involves the use of keying elements which are used in a repetitive manner and which, in a single message, operate in conjunction with a fixed number of cipher alphabets, as would the example shown in fig. 5.

Polyalphabetic systems of the aperiodic type are of great diversity in construction. In these systems periodicity is either entirely lacking because of the nature of the method or it is suppressed by incorporating elements which serve as periodicity interrupters. In the well-known aperiodic system designated as the running-key or nonrepeating-key system, periodicity is avoided by employing for the key the running text of a book, identical copies of which are in possession of the correspondents. The starting point of the keying sequence is agreed upon in advance,

Key:	T R O Y T	R O Y T R	O Y T R O	Y T R O Y
Plain-text:	L E T T E	R N O T Y	E T R E C	E I V E D
Cipher:	E V H R X	I B M M P	S R K V Q	C B M S B

Fig. 5.—An example of multiple alphabet substitution. The alphabets of the simple Vigenère table (fig. 4) are used in conjunction with the keyword TROY, which is written above the plain text and repeated as many times as required. Each letter is then enciphered by the keyletter above it. Thus, for the first letter, $L_p(T_k) = E_c$; for the second letter, $E_p(R_k) = V_c$, etc. Usually all the plain-text letters governed by the first letter of the key are enciphered in succession, then those governed by the second, etc., this method reducing the possibility of error. In practice, the keyword, and therefore the period, is usually longer than that employed in this example, words and phrases of 15 to 20 or more letters being common. Much more important, however, is the fact that the alphabets employed in modern times are seldom those of the simple Vigenère table, but are mixed alphabets, the use of which makes the cryptograms more secure against solution.

or is indicated in one of many ways by means of elements called indicators, which designate the exact point where the sequence begins. Another type of nonrepeating-key system, often designated as the autokey system, is that in which, beginning with a preconcerted key letter, each cipher letter (or sometimes each plain-text letter) becomes the key letter for the encipherment of the next plain-text letter.

The various systems of substitution thus far mentioned are monographic in nature, involving single plain-text letters. In digraphic systems substitution is by pairs of letters and uses preagreed cipher squares or matrices in which the letters of the alphabet are disposed according to some key and the cipher equivalents of pairs of plain-text letters are found by following prearranged rules. One of the best-known systems of this sort is that called the Playfair cipher, named after Lyon, Baron Playfair, but invented by Sir Charles Wheatstone. It is illustrated in fig. 6.

Properly selected methods of transposition and substitution when combined into a single system result in producing cryptograms of great security, but because of the additional complexities of operation introduced by the combination, with the resultant increased possibilities of error, such systems are not often encountered in practice.

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

1. Message: AM FOLLOWINQ ON NEXT TRAIN.
2. Plain-text: AM FO LX LO WI NG ON NE XT TR AI NX
3. Cipher: GR QF EI XG YX LB BX LX EO AP HA LI

Cryptogram

4. GRQFE IXGYX LBBXL XEOAP HALI

Fig. 6.—The Playfair cipher; an example of digraphic substitution. An alphabet square of 25 letters (I serving also as J) based upon a keyword (in this case "Lexington") is drawn up, and substitution is performed with pairs of letters. Three cases as regards the position occupied in the square by the members of a pair are possible; they lie (1) in the same horizontal row, or (2) in the same vertical column, or (3) at the opposite ends of one of the diagonals of an imaginary rectangle. By studying the following examples under each case the method of finding equivalents will become apparent: Case (1) LI = EN, AB = BG, OB = AG; Case (2) ED = TP, EV = TE, PV = VE; Case (3) ER = IP, IM = LR, ST = PB. The text of the message to be enciphered, line 1 in the figure, is divided up into pairs, line 2. Double-letter pairs in the text are to be separated by an interposed letter such as X, as shown in the case of the word "following" in this example. Note also the addition of the letter "X" to the end of the message in order to make a pair of letters. The cipher equivalents are shown in line 3, the final cryptogram in five-letter groups, in line 4. Various arrangements of letters in the alphabet square are possible, and completely disarranged alphabets may be employed.

Code Systems.—Basically, a code system is merely a type of substitution system in which the textual units that undergo cryptographic treatment are of varying and unequal lengths. Generally they are entire words, but often they may be single letters, syllables, numbers or phrases of varying length and sometimes whole sentences. The cryptoprinciple underlying code systems is very old and has as its essential element a specially compiled list of words, phrases and sentences, each of which is accompanied by a code group which may be another word (real or artificial) or a group of letters or figures. Identical copies of specialized lists of the foregoing sort, called codebooks or codes, are held by the communicators. In encoding a message it is necessary merely to refer to the codebook and replace the words, phrases or sentences composing the message by the code groups assigned them, constructing the code message, step by step, in this manner and using the fewest code groups possible. As to their cryptographic construction or arrangement of contents, codes may be of the one-part or the two-part type, the principal difference between them being shown in fig. 7.

The condensing power of a code obviously depends upon the extent and the nature of its vocabulary. The security of a code system depends somewhat upon the distribution and construction of the codebook, but mostly upon whether it is used in conjunction with a good cipher system which is superimposed on the code text and the purpose of which is to afford secrecy, in the case of purchasable or publicly available codes, or additional secrecy, in

One-part Code	Two-part Code	
	Encoding	*Decoding*
ABABA—A		
ABACE—Abandon-ing-s	KABOL—A	ABABA—Abeyance
ABADI—Abandoned	STOLG—Abandon-ing-s	ABACE—Procedure
ABAFO—Abate-ing-s	EXIFO—Abandoned	ABADI—To purchase
ABAGU—Abated	ZUMRA—Abated	ABAFO—Commenced
ABAHY—Abeyance	ABABA—Abeyance	ABAGU—Do not think
ABEBE—Abide-ing-s	ROABY—Abide-ing-s	ABAHY—Recorded
ABECI—Abided		
ZYZYZ—Zone-s	BIKUR—Zone-s	ZYZYZ—According to

Fig. 7. —Extracts from typical one-part and two-part codes. In the one-part type the code groups and the vocabulary are arranged in parallel, alphabetic (or numerical) sequences, so that a single book serves for encoding as well as for decoding. In the two-part type the encoding book lists the elements of the vocabulary in alphabetic order but the code groups are in random order, so that a decoding book, in which the code groups appear in alphabetic (or numerical) order accompanied by their meanings, is essential. The degree of secrecy afforded by a code of the latter type is much greater than that afforded by one of the former type, all other things being equal.

the case of private or governmental codes. Code messages which undergo this second step are said to be superenciphered, reenciphered or, simply, reciphered. The principal purpose of code in commercial communications is to effect economy in their cost of electrical transmission, secrecy being usually of secondary importance (except in certain types of banking operations). Codes for general business communications are purchasable from their publishers and therefore in themselves provide no secrecy. Many business firms, however, use specially compiled private codes which, if carefully restricted in distribution, may be regarded as confidential or secret. In governmental and especially in diplomatic or military communications secrecy is generally of primary importance, economy is secondary.

Governmental codes which are intended to be secret are, of course, very carefully guarded in their production, distribution and usage and, as a general rule, messages in such codes are superenciphered.

Cryptoapparatus.—Cryptodevices and cryptomachines vary in complexity from simple, superimposed, concentrically or eccentrically rotating disks to large mechanical machines and electrically operated cryptoteleprinting apparatus. One of the best devices of the more simple, mechanical type is that known as the Bazeries cylinder (*see* fig. 8), named after the French cryptographer who is commonly credited with its invention in 1891. The principle upon which the device is based was, however, conceived

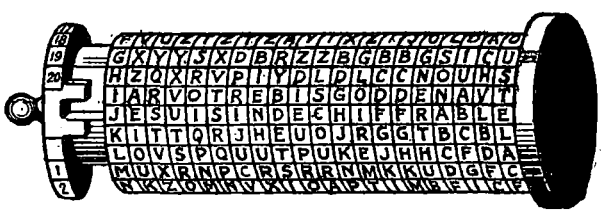

FIG. 8.—THE BAZERIES CYLINDER

The Bazeries cylinder consists of a set of 20 disks, each bearing on its periphery a differently mixed alphabet. The disks, which bear identifying numbers from 1 to 20, are assembled upon the shaft in an order that corresponds to a numerical key. To encipher the message, 20 letters are taken at a time, and the disks are revolved so as to align the 20 letters horizontally; the letters of any row can be taken for the cipher text. In deciphering, the cipher letters, taken 20 at a time, are aligned horizontally; the disks are then locked into position. By slowly revolving the whole cylinder and examining each row of letters, one and only one row will be found to yield intelligible text all the way across

many years before by Thomas Jefferson (*see* Jefferson's *Papers* in the Library of Congress, vol. 232, item 41,575).

Devices of the sort exemplified in fig. 8 soon proved to be inadequate for modern cryptocommunications, as regards not only speed, accuracy and facility in operation but also security of the end product. It was not long, therefore, before more automatic

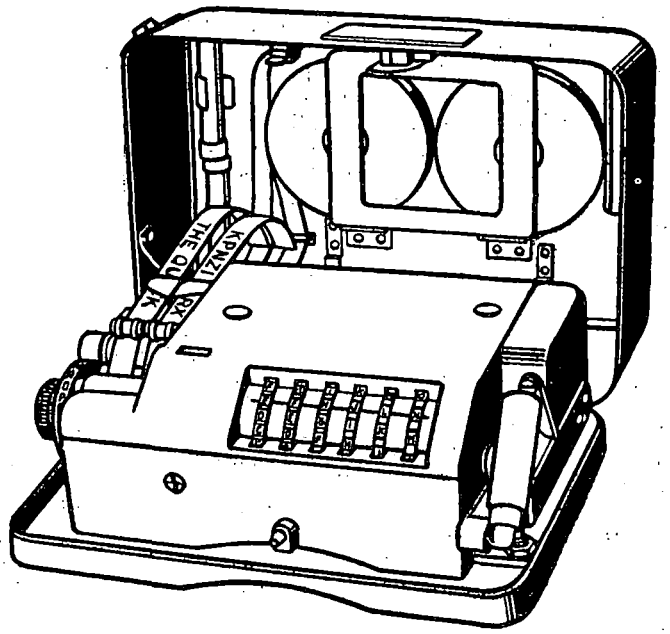

FIG. 9.—PORTABLE CRYPTODEVICE WHICH PRODUCES A PRINTED RECORD OF THE CIPHER TEXT AS WELL AS OF THE PLAIN TEXT. THE LETTERS TO BE ENCIPHERED OR DECIPHERED ARE SET BY TWIRLING THE INDICATING DISK SHOWN AT LEFT AND OPERATING THE LEVER AT THE RIGHT. THE SIX WHEELS CONTROL THE CRYPTOGRAPHY, IN CONJUNCTION WITH CERTAIN OTHER VARIABLE KEYING ELEMENTS INSIDE THE MACHINE

and more secure types of apparatus were invented and developed. In such apparatus rotatory components referred to as cipher rotors have come to be of primary importance. The rotors may be of a mechanical or an electrical type. Fig. 9 shows a machine which uses a series of mechanical rotors to produce an extremely long, continuously changing key for encipherment. Although the results of manipulating the machine are printed upon a paper tape, the absence of a typewriter keyboard makes operation of the machine slow and tedious. Fig. 10 shows a machine which uses a series of juxtaposed electrical rotors and stators to form a path for the passage of electric currents connecting the 26 keys of a keyboard to the 26 lamps of a light board upon which the results of manipulating the keys are indicated. Automatic angular displacements of the rotors serve to vary the path with each depression of any key of the keyboard. From a practical point of view such a machine is not satisfactory for offices engaging extensively in cryptocommunication since it lacks an automatic recording or

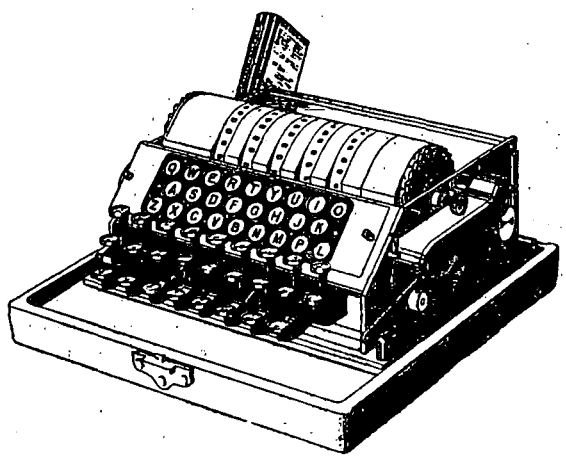

FIG. 10.—ELECTRICAL CIPHER MACHINE EMPLOYING A TYPEWRITER KEYBOARD

printing mechanism and the results of operating the keyboard have to be recorded by the operator by hand. Therefore, a machine combining both a keyboard and a printing mechanism had to come, sooner or later. Such a combination of components is shown in

fig. 11. Machines of this type, which are not necessarily associated with the electrical communication system but which merely produce an end product (a printed record) that can be given to a communication centre or telegraph office for transmission, are called off-line cryptomachines.

But even this stage in improvement in cryptocommunications

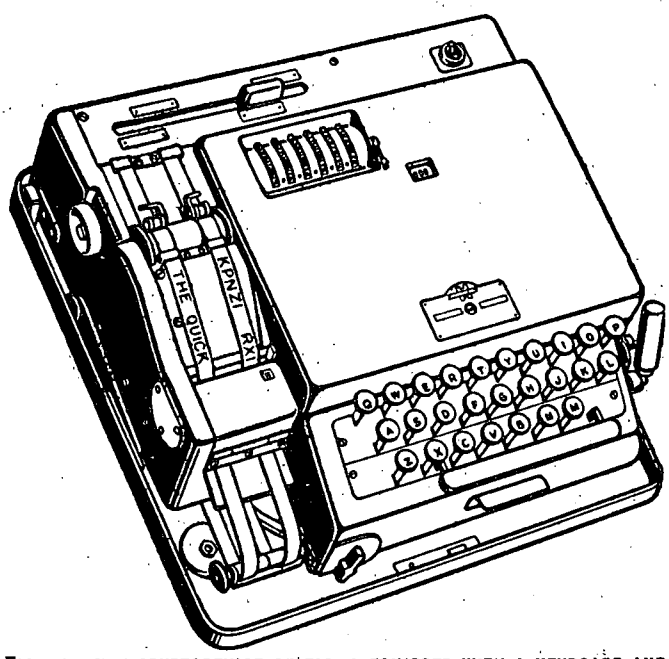

FIG. 11.—THE CRYPTODEVICE OF FIG. 9 EQUIPPED WITH A KEYBOARD AND DRIVEN BY AN ELECTRIC MOTOR. THE TWO MACHINES ARE CRYPTOGRAPHICALLY INTERCOMMUNICABLE. THE OPERATING LEVER AT RIGHT IS USED IN CASE OF POWER FAILURE

proved inadequate and communicators soon began to see need for on-line cryptomachines, *i.e.*, apparatus which combines in a single and instantaneous operation the following steps: (1) manipulation of a key of the keyboard at the transmitting station to correspond to a character of the plain-text message; (2) automatic encipherment of that character to form a cipher character; (3) electrical representation of that character by a signal or a permutation or combination of signals corresponding to the cipher character; (4) electrical transmission of the signal or signals; (5) their reception at the distant end; (6) their translation into a cipher character; (7) automatic decipherment of the character; and (8) printing the deciphered character—all this at the rate of at least 300 characters (=60 words) per minute. In fig. 12 is shown a prototype of such a machine. It is obvious, of course, that machines of this advanced type can be employed only where there is direct access to transmission and reception facilities.

As a result of extensive research and development after 1920, machines of the sorts here described have undergone considerable improvement, and further progress in written-cryptocommunications technology appears to lie in this direction.

Signal Intelligence.—The principal components of signal intelligence are: (1) communication intelligence, derived from the interception and analysis of signals which are involved in the exchange of messages or communications between persons and which are, therefore, of the type designated as communication signals; and (2) electronic intelligence, derived from the interception and analysis of electromagnetic radiations which are of a type other than those used in communications, such as radar, identification and recognition signals, navigational beacons, etc., and which are therefore designated as noncommunication signals. It should be noted, however, that it is often difficult to draw a line of demarcation between communication intelligence and electronic intelligence because these two fields deal with signals which merge into each other in the continuous spectrum generally referred to as that pertaining to communications electronics.

Communication Intelligence.—This phase of cryptology deals with the processes, methods or means employed in deriving information by intercepting and analyzing enemy communications. Its principal components are: (1) interception and forwarding of traffic (messages) to analysis centres; (2) traffic analysis, including radio direction or position finding and operator identification; and (3) cryptanalysis or solution (and translation, when necessary) of the texts of the messages. Only the last two components will be discussed.

Traffic Analysis.—Stated in general terms, traffic analysis involves studying the messages exchanged within a communications network for the purpose of penetrating the signal security camouflage superimposed thereon.

Such study permits reconstructing the networks from data such as volume, direction and routing of messages, frequencies, schedules and call signs used, etc. When the most important features of the networks have been thus ascertained, the analyst is not only able to ascertain the geographic locations and dispositions of military units (order of battle) and important movements of these units, but also to predict with varying degrees of reliability (based upon inferences) the area and extent of future military tactical or strategic operations.

Cryptanalysis.—The science of solving cryptograms by analysis is called cryptanalysis, to distinguish the indirect methods of reading cryptograms from the direct methods which, of course, require a knowledge of the basic method and specific key, in the case of ciphers, or possession of the codebook, in the case of codes. Apart from the more simple, classical types, nearly every scientifically constructed cryptographic system presents a unique case in cryptanalysis, the unravelling of which requires the exercise of unusual powers of observation, inductive and deductive reasoning, much concentration, perseverance and a vivid imagination; but all these qualities are of little avail without a special aptitude arising from extensive practical experience. It is worthwhile to note that the resistance which a specific cryptosystem will have against cryptanalysis is often vitally affected by the sophistication of the rules for its use and by the degree to which these rules are observed by cipher clerks. In respect to the latter, Francis Bacon's comment in his *Of the Advancement of Learning* (1605) is as true today as the day it was written: "But in regarde of the rawnesse and unskillfulnesse of the hands, through which they passe, the greatest Matters, are many times carryed in the weakest cyphars."

A preliminary requisite to the analysis of a cryptogram is a determination of the language in which its plain text is written, information which is either already at hand in the case of official communications or which, in the case of private ones, can usually be obtained from extraneous circumstances. Next comes a determination as to whether a cipher or a code system is involved;

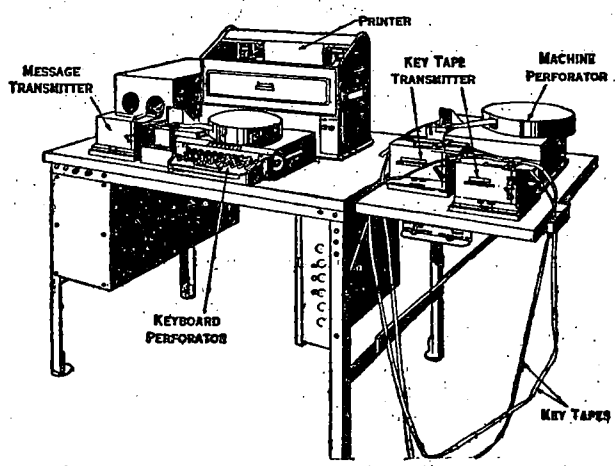

FIG. 12.—A PRINTING TELEGRAPH CIPHER MACHINE

this is based upon the fact that differences in their external appearance are usually sufficiently well marked to be detectable. If the cryptogram is in cipher, the next step is to determine whether transposition or substitution is involved. This determination is

made on the basis of the fact that in plain text the vowels and consonants are present in definite proportions. Since transposition involves only a rearrangement of the original letters, it follows that if a cryptogram contains vowels and consonants in the proportions normally found in plain text in the language in question, it is of the transposition class; if not, it is of the substitution class. The solution of transposition ciphers involves much experimentation with matrices of various types and dimensions, clues to which are afforded by the number of letters in the messages and extraneous circumstances. The assumption of the presence of probable words is often necessary. Special methods of solution based upon a study of messages of identical lengths, or with identical beginnings or endings, are often possible to apply when much traffic has been intercepted. Finally, the presence of letters which individually are of low frequency but which when present have a great affinity for each other and form pairs of moderate or high frequency, such as *qu* in Spanish or *ch* in German, afford clues leading to solution.

The basis upon which the solution of practically all substitution ciphers rests is the well-known fact that every written alphabetic language manifests a high degree of constancy in the relative frequencies with which its individual letters and combinations of letters are employed. For example, English telegraphic texts show the following relative frequencies in 1,000 letters, based upon an actual count of 100,000 letters appearing in a large but miscellaneous assortment of telegrams of a commercial and governmental nature, but all in plain language:

E	T	R	I	N	O	A	S	D	L	C	H	F
126	90	83	76	76	74	72	68	40	38	33	33	30
U	P	M	Y	G	W	V	B	X	K	Q	J	Z
30	27	25	21	18	14	13	11	5	3	3	2	1

These characteristic relative frequencies serve as a basis for identifying the plain-text values of the cipher letters, but only when the cipher has been reduced to its simplest terms. Thus, the problem of solving a monoalphabetic substitution cipher involves only one step, since the text is already in the simplest possible terms, and regardless of the kind of cipher alphabet employed practically every example of 25 or more characters representing the monoalphabetic encipherment of a "sensible" message in English can readily be solved by the well-known principles of frequency, made popular by Edgar Allan Poe's romantic tale *The Gold Bug*. However, the problem of solving a polyalphabetic substitution cipher involves three principal steps: (1) determining the number of cipher alphabets involved; (2) distributing the cipher letters into the respective individual frequency tables to which they belong; and (3) analyzing each of the latter on the basis of normal frequencies in plain text of the language involved. In the case of aperiodic ciphers, because of the absence of cyclic phenomena, these steps are often very difficult, especially when the volume of text is limited. Frequently the only recourse is to employ repetitions as a basis for superimposing separate messages so that, irrespective of the number of alphabets involved or their sequence, the letters pertaining to identical cipher alphabets fall into the same columns, and then the respective columns are treated as monoalphabetic frequency tables. The analysis of the frequency distributions of a polyalphabetic cipher is effected much more readily when the alphabets are interrelated than when they are independent.

The question as to whether an absolutely unsolvable cipher system can be devised is of more interest to laymen than to professional cryptographers. Edgar Allan Poe's dictum that "it may be roundly asserted that human ingenuity cannot concoct a cipher which human ingenuity cannot resolve" is misleading unless qualified by restricting its application to the great majority of the practical systems employed for a voluminous, regular correspondence. Isolated short cryptograms prepared by certain methods may resist solution indefinitely; and a letter-for-letter cipher system which employs, once and only once, a keying sequence composed of characters or elements in a random and entirely unpredictable sequence may be considered holocryptic, that is, messages in such a system cannot be read by indirect processes involving cryptanalysis, but only by direct processes involving possession of the key or keys, obtained either legitimately, by virtue of being among the intended communicators, or by stealth.

History.—It may as well be stated at the outset that as of the 1950s there did not exist in any language a detailed, authentic and publicly available history of cryptology. Moreover, because of the curtain of secrecy which is invariably placed around cryptologic work of an official character, accurate accounts of historically important events or of noteworthy inventions and improvements in cryptologic technology usually enter into the public domain only many years after the event or invention has occurred. Therefore, although it is difficult or impossible to ascertain with certainty much about the origin of any specific item or fact of cryptographic or cryptanalytic importance, the data should be traced back at least as far as the open or public records will permit. With this limitation in mind, this account will begin by noting that secret modes of signalling and communication have probably been in use from the earliest times, since the desire or need for secrecy in communication is certainly as old or nearly as old as the art of writing itself. However, mysteries such as the prophetic and apocalyptic writings of the orient and the sayings of the Sibylline oracles are generally not regarded by cryptologists as coming within their province; nor do they generally do anything more than merely refer to the various systems of stenography or shorthand used since the time of the Romans, including that known as Tironian notes, named after Tullius Tiro, Cicero's learned freedman and friend, who elaborated a system which was popular for almost 1,000 years. Although it is valid to assume that cryptography was used by all of the peoples of antiquity, the assumption has been confirmed in the case of Egyptian hieroglyphic writing by the outcome of studies which were begun by Jean François Champollion himself, were continued sporadically by other students for many years and culminated in 1932, disclosing that not one but three different sorts or degrees of cryptography were actually used by the ancient Egyptians. Cryptography was practised among the ancient Jews, whose Talmudic scholars dealt with it as a part or phase of their Kabbalah, which includes certain operations of a cryptographic nature. The ancient Greeks were users of the art, and at least one cryptographic device, called the scytale, is known to have been employed by the Lacedemonians for secret communications between military commanders in the field and their superiors at home. In the writings of Aeneas Tacticus (360–390 B.C.) is to be found the very earliest treatise on cryptography thus far discovered; in addition to treating of secret dispatches, they contain a description of a cipher disk, a detailed explanation of which cannot here be included. Numerous ancient Greek documents exist which are either partially or wholly in cipher; and one cipher alphabet found therein has been traced back to the 9th century A.D. Despite the fact that Roman cryptographic documents are rather rare it is well known that the Romans used ciphers, for there are numerous references to those which Caesar and Augustus employed. In fact, the name Caesar is often used in cryptologic literature to designate the type of cipher alphabet and cipher system known as the monoalphabetic cipher using standard alphabets. In the middle ages (*c.* 450–1450) cryptography was employed rather infrequently, for the most part in connection with the pseudo sciences of alchemy and astrology. The most widespread cryptographic system of that period corresponds to that used by the Roman emperor Augustus, in which a letter was merely replaced by the one following it in the normal alphabet (but *z* was replaced by *aa*). Most often, however, only certain letters were thus replaced, sometimes only the vowels.

The beginnings of modern cryptography can be traced back to Italy, the birthplace, as well, of modern diplomacy. Many cipher alphabets have come down to us which were used in the official messages of the papacy as well as in those of the early Italian republics. Ciphers were used in Venice as far back as in 1226; in Mantua and in Modena as early as 1305; and in Lucca, Florence, Siena, Pisa and Milan before 1450. It was also in Italy, beginning soon after 1500, that cryptologic operations first came to be organized on an effective basis. The invention of new cipher systems was promoted and reserves of cryptographic vocabularies

were prepared and kept in readiness for prompt issue as replacements for old or compromised ones. The earliest extant piece of work of a cryptographic nature is a small manual or compilation of the ciphers used about 1379 by Gabriel de Lavinde of Parma and preserved in the Vatican archives. In these ciphers all the letters were represented by arbitrary symbols and the vowels were treated no differently than the consonants in regard to the number of equivalents assigned them. However, some of these ciphers included nulls and others included brief lists of words and proper nouns, compilations first called nomenclators, then repertories and, later, small codes. By 1400 A.D. it had become apparent that each of the vowels should have more than a single cipher equivalent and there is a record of a cipher system involving a reversed standard alphabet with three different supplementary symbols as variants for each vowel. By the 15th century, according to Aloys Meister, Italian cryptography had been elaborated to the point where "three to six different symbols could be used to represent a single letter of the alphabet, the individual syllables—arranged alphabetically for this purpose, ba, be, bi, bo, bu, ca, ce, ci, etc.,— had specific cipher equivalents, and an ever-increasing number of complete words were incorporated into the nomenclators. Their abundance in content became so great that it was possible to fill a lengthy alphabetic index with the special equivalents for syllables and words . . ." The first complete cipher, *i.e.,* one containing arbitrary symbols for each of the letters and variants for the vowels, as well as nulls and a nomenclator, is exemplified by a Venetian cipher of 1411, cited by Luigi Pasini. In the first half of the 15th century, after some further expansion in content, nomenclators attained such a state of development and practical utility that they remained for centuries the prototype of the diplomatic repertories used by nearly all European governments, as well as by that of the young and rapidly growing republic in the western hemisphere, the United States. Before continuing with the account of the further development of cryptosystems of this sort it may be well to direct attention to the invention and development of another cryptosystem which, originating in the very simple devices of the ancients, culminated some time during the 16th century in the so-called *chiffre indéchiffrable* ("the indecipherable cipher") often referred to as the Vigenère cipher (fig. 4). It should be noted, however, that Vigenère's description of the cipher differs decidedly from the form usually ascribed to him and presents an essentially more difficult problem to the cryptanalyst, that Vigenère nowhere speaks of the cipher as the *chiffre indéchiffrable par excellence,* and that, except for an autokeying principle, Vigenère lays no claim to having originated the cipher.

Despite the much-advertised virtues and security of the Vigenère cipher, practical cryptographers and cryptanalysts tended to put their faith rather in the older nomenclators and repertories, and it is therefore necessary to turn once more to these cryptoaids and follow their progress.

The repertories used by France in the 16th century, under Louis XIII and Louis XIV, underwent important improvement with the introduction of an innovation which is now called the randomized or two-part arrangement of contents, illustrated in fig. 7. This improvement also soon found its way into the official cryptography of England and other countries. The repertories used by the papal court in the same century incorporated an additional new feature, that of making some or all of the cipher characters represent two or more different letters. After attaining this fairly advanced state of development, European cryptography went into a decline that reached its lowest level under Napoleon I; it is possible that one of the factors leading to the disaster which overtook him in Russia was the solution by the Russians of intercepted French ciphers. After the middle of the 19th century, stimulated perhaps by the spreading use of electromagnetic telegraphy, there came an expansion in the content of repertories, soon to be called codes. By the end of that century large codes containing 100,000 or more words and phrases were compiled not only for governmental but also for commercial communications. In such codes, of course, the length of the code equivalents had to be increased too, and code groups came to be composed, first, of groups of figures or of bona fide dictionary words, then of artificial words and, later, of five-letter groups constructed scientifically so as to obtain the maximum not only in economy in cost of transmission but also in efficiency in the correction of transmission errors. For another 75 years, beginning about 1860, the cryptoprinciples embodied in the early nomenclators but now expanded into large codes were the ones preferred for diplomatic and commercial cryptocommunications; literal or letter-for-letter ciphers were used only rarely for such communications. On the other hand, for military cryptocommunications, cipher systems of the latter type were preferred, except for high-level or strategic communications. In the U.S., during the War Between the States (1861–65), the Federal army employed small repertories in connection with word transposition, the so-called route cipher. The Confederate army used the Vigenère cipher—which the Federal army cryptanalysts are said to have solved every time a message in it was intercepted.

During the first two years of World War I cipher systems were used almost to the exclusion of code systems by all belligerents for protecting tactical communications in the field of operations, although code systems continued to be used for diplomatic and high-level strategic military cryptocommunications. By 1917, however, codes came to be used for the secret communications of the smaller and intermediate-size military formations. The extent of their vocabularies varied with the size of the unit, so that the code for communications between large units might contain 10,000 words and phrases, whereas that for communications between small units might contain only a few hundred or less. After about 1925 the direction of development and improvement in governmental cryptocommunications tended to explore and exploit the possibilities of automatic cipher machines, as indicated earlier in this article, and this took place in all types of secret communications: diplomatic, military, naval, air, etc. This change in direction of evolution and development of cryptosystems, code systems giving way to cipher systems in practical cryptocommunications where secrecy is the primary consideration, is only an obvious result, in the field of communications, of the increased tempo of mechanization in all fields since the beginning of the 20th century. Cipher systems, the units of which are generally single letters, lend themselves much more readily to mechanization than do code systems, the units of which are generally complete words and often long phrases or sentences, because the number and lengths of different permutations and combinations of electrical signals needed to represent the relatively small number of different basic units of a written alphabetic language (in English 26 letters) are very much smaller than the number and lengths of different permutations and combinations that would be required to represent the large number of different words in such a language, not to mention phrases and sentences.

This brief account of technological developments in cryptography has its counterpart in cryptanalysis, but if it is difficult to ascertain with certainty data concerning the origin of any specific cryptosystem, because of the veil of secrecy already mentioned, it is almost impossible to ascertain with certainty by whom, where or when a specific cryptanalytic principle or process was first conceived or employed. The secrecy veil in this area becomes an almost impenetrable curtain, so that it is certainly valid to assume that news of cryptanalytic success or of the invention of a new technique becomes public only many years after the events or, perhaps, never.

Under these circumstances, one hardly expects to find a historical account of the very first success in cryptanalysis, but the earliest brochure on record dealing with cryptanalytic theory is that of the Leone Battista Alberti, whose *Trattati in cifra,* written between 1467 and 1472, deals not only with cryptanalytic theories and processes but also with cryptography and statistical data. It is, therefore, the oldest treatise on cryptology in existence. The brief treatise of Sicco Simonetta, a cryptanalyst of the Sforzas at the court of Milan, which is dated July 4, 1474, and which is a strictly practical guide to cryptanalytic procedures, is the oldest treatise in the world purely devoted to cryptanalysis. Beginning

about 1506 and until 1539 another Italian professional cryptanalyst, Giovanni Soro, devoted his leisure hours to the preparation of a treatise which has never been found. As was the case with cryptography, so on the cryptanalytic side the activities in the various small Italian republics and the papacy came to be organized as serious enterprises. In order to protect these activities against indiscretions or "leakage," the principles of secrecy and security were firmly established and wilful violations were treated as capital offenses. The development of technical skill was rewarded, and instruction in cryptanalysis was often carried on from one generation to the next in the same family. But, as in the cryptographic field, so in the cryptanalytic field there have been periods of progression and retrogression in technology. However, after about 1930 cryptanalytic technology progressed rather notably. The mechanization which affected cryptography also affected cryptanalysis, and the invention and development or improvement of machines to aid in or speed up the tedious processes usually involved in the solution of complex cryptosystems exercised an important effect upon the standards by which cryptosecurity had to be measured.

The foregoing is a brief history of technological developments in cryptology. Nothing has yet been said about the usually very secret but quite important roles which poor cryptography or good cryptanalysis have played in international relations, both in peace and war, and in respect to both diplomatic and military affairs. As already stated, the historical accounts of these instances usually lag many years behind the events to which they relate. But in at least one case in modern history the revelation of the spectacular role played by cryptology in diplomacy and warfare took place not very long after the event. Hearings held in Washington, D.C., 1945–46, by the joint committee on the investigation of the Pearl Harbor attack, made it clear that shortly before and during World War II U.S. cryptanalysts had considerable success in solving Japanese codes and ciphers, success indeed to the extent that it was possible for the committee to state that all witnesses familiar with the intelligence produced "have testified that it contributed enormously to the defeat of the enemy, greatly shortened the war, and saved many thousands of lives." For other instances the reader must consult some of the books listed in the bibliography, but not many will be found that pertain to relatively recent history.

BIBLIOGRAPHY.—Although no comprehensive and authentic history of cryptology had been published by the 1950s, it is interesting to find that there does exist a comprehensive and scholarly bibliography of cryptologic literature covering practically all publicly available books and important papers on the subject: Joseph S. Galland, *An Historical and Analytical Bibliography of the Literature of Cryptology* (Evanston, Ill., 1945). Such a compilation is more helpful to specialists than to the general reader, therefore, a selection of items from the Galland bibliography may be useful. Also, after 1945 there were published a few items of importance which should be mentioned herein. Modern works worthy of special mention are as follows: André Lange and E.-A. Soudart, *Traité de cryptographie* (1925); Marcel Givierge, *Cours de cryptographie* (1925); Roger Baudouin, *Éléments de cryptographie* (1939); H. F. Gaines, *Elementary Cryptanalysis* (Boston, 1939; London, 1940); Luigi Sacco, *Manuale de crittografia,* 3rd ed. (1947); Charles Eyraud, *Precis de cryptographie moderne* (1953). (W. F. F.)