

PRINTED LECTURES  
I, II, III, IV  
(Nos V & VI not received)

~~CONFIDENTIAL~~

NSA

TECHNICAL  
JOURNAL

VOL. VI

WINTER 1961

NO. 1

	Page
Remarks at the Dedication of von Neumann Hall .....	1
The Association Factor in Information Retrieval .....	
H. EDMUND STILES	7
Introduction to Cryptology—IV .....	
WILLIAM F. FRIEDMAN	25
Pattern Recognition .....	
WALTER W. JACOBS	77

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~

# THE NSA TECHNICAL JOURNAL

## EDITORIAL POLICY BOARD

S. KULLBACK, *Chairman*

WILLIAM A. BLANKINSHIP	ARTHUR F. MATHISEN, <i>Sec. Advisor</i>
LAMBROS D. CALLIMAHOS	IGNATIUS G. MATTINGLY, <i>Exec. Sec'y</i>
HOWARD H. CAMPAIGNE	HAROLD J. STUKEY
DANIEL M. DRIBIN	COL. M. J. BARTOSIK
PAUL F. FRIEDMANN	WILLIAM D. WRAY

## EDITORIAL STAFF

EXECUTIVE EDITOR

PAUL F. FRIEDMANN

PANEL OF EDITORS

LAMBROS D. CALLIMAHOS, *Chairman*

JOSEPH BLUM	LOWELL K. FRAZER
GEORGE L. CHESNUT	FRANK W. LEWIS
MORTON KUPPERMAN	CHARLES W. RECHENBACH
EDWIN C. FISHEL	C. RUSSELL SUMMERS

*NOTICE: This material contains information affecting the National Defense of the United States within the meaning of the Espionage Laws, Title 18, U. S. C., Sections 793 and 794, the transmission or the revelation of which in any manner to an unauthorized person is prohibited by law.*

The *NSA Technical Journal* is published four times a year under the direction of the *NSA Technical Journal* Editorial Policy Board. Telephone: OUTSIDE: 7249, SECURE: 3057. Any cleared and indoctrinated person may be permitted access to the *Journal* by a regular receiver of the *Journal*, or by the Library. Copies of the *Journal* which are no longer required may be destroyed, and the accompanying certificate of destruction filled out and returned to the Office of Administrative Services.

NSAL—S-129,098

Use of funds for printing this publication approved by the Director of the Bureau of the Budget, 20 January 1960.

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~

## Extending the Challenge

We and our customers are continually involved in evaluating the two most significant aspects of COMSEC systems—security and logistics. Singly or in combination they present a fascinating challenge for devising the means to facilitate reaching rational decisions which provide the best over-all balance. Despite our efforts there is still no formula, no clear-cut criteria for weighing the associated variables. The implications of a wrong “guess” are shattering in terms of impact on our national security. Dozens of parallel examples could be cited for SIGINT.

While the activities of the Agency continue to mushroom in complexity, it is important that our approaches to problems keep pace.

There is within our grasp the essence of a solution. It is no accident that we have engaged in some operations research, applying our “cryptologic” talents in statistics, mathematics, and engineering to what are actually “management” problems. The very fact that our in-house scientific skills can be blended to analyze a “cryptologic-management” question holds promise of possibilities and hope.

The cryptologist-manager of tomorrow must acquire system and discipline needed for sound planning; but he has not yet learned how to judge problems in all the necessary dimensions, to establish a balance, and to decide things not on intuition alone, but on a sounder more scientific base.

What we lack is a way to treat with assurance typical situations which require that the solution to the “flap” be melded with the needs of the future; that our capabilities project the demands of our customers; and that individual readjustments not threaten the Agency as a whole or any of its missions. This will be possible only through the use of a logical structure which provides definitions, specifications, measurements, and a common communications medium.

We are lucky that cryptology and management are both infants as professions and as sciences. There may be likenesses in their individual, general methodologies, some possibility of a harmonizing logic. A common, symbolic, cryptologic language is beginning to emerge. The stimulating analogy is that the management language is headed toward a similar integrated and synthesized structure! Further, it could become complicated by moral and social considerations.

---

Guest Editor for this issue is Mr. Paul E. Neff, Assistant Director NSA for Communications Security.

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~

The cryptologic skills, problems, and needs we have today offer a fabulous opportunity to experiment more fully in the management area; and the possibilities fire the imagination.

## CORRECTION

In the July-October 1960 issue of *The Journal*, Dr. A. Sinkov was erroneously referred to as Assistant Director of PROD, NSA. Dr. Sinkov's correct title is Deputy Director, PROD, National Security Agency.

~~CONFIDENTIAL~~

UNCLASSIFIED

## Remarks at the Dedication of John von Neumann Hall

*Unclassified*

*A new building to house the NSA Research Institute was dedicated on October 22, 1960 at Princeton University. The major dedication addresses were delivered by Dr. James R. Killian, Jr., Chairman of the Board, Massachusetts Institute of Technology, and Dr. Robert Francis Goheen, President of Princeton. The addresses are reproduced here with introductory remarks by Dr. Howard H. Campaigne, Chief, Office of Research, NSA.*

The Institute for Defense Analyses (IDA) is a non-profit organization which was founded in 1956 by five of the leading research institutions in our country: Massachusetts Institute of Technology, Tulane University, California Institute of Technology, Case Institute of Technology and Stanford University. It has one broad purpose—to carry out research on problems referred to it by the Defense establishment.

Two years later, the Baker Committee proposed that a separate research facility be established, preferably in an atmosphere conducive to intellectual endeavor. The NSA Scientific Advisory Board, of which Professor von Neumann of Princeton was a member, strongly endorsed the concept. The recommendation was taken to the White House.

President Eisenhower approved the establishment of an activity to perform independent cryptologic research in February of 1958 and the search for a suitable location was underway. Among the groups approached were Syracuse University (with the possible help of General Electric), The Rand Corporation, IBM (with Stanford University), Harvard University, Princeton University and, of course IDA.

IDA undertook the project. It immediately set up a new division, the NSA Research Institute (also known as Focus), to do the work and entered into a contract with Princeton for a new building to house the effort. It is the dedication of this building which was the occasion of the following speeches.

*Dr. J. R. Killian, Jr. . .*

I speak in behalf of President Norton, the trustees and staff of the Institute for Defense Analyses in expressing appreciation and satis-

UNCLASSIFIED

## UNCLASSIFIED VON NEUMANN HALL DEDICATION

faction in the splendid opportunities and facilities provided by Princeton University for the work of the Institute's Communications Research Division. We are especially proud and grateful that the University has made it possible to have this fine new building available for the Institute's communications research program and that it has been possible for us to join with the University in making this building a memorial to Dr. John von Neumann. Because of his great accomplishments as a mathematician and his remarkable contributions to the public service of his adopted country, it seems happily appropriate that this building have the distinction of his name.

From the standpoint of the Institute for Defense Analyses, this whole enterprise here in Princeton has been a happy undertaking. The arrangements entered into with the University for bringing the building into being have worked well. We have welcomed, too, the opportunity to join with Princeton in planning for the design and use of the computer, one of the facilities housed in the building. We are glad that the University community has access to this fast and versatile computer and that other parts of the program managed by IDA here have been so arranged as to permit a fruitful relationship between the Communications Division and the scholarly community. This is possible because a part of the work of IDA's Communications Division is in fundamental areas of mathematics and associated communications sciences, which are suited to university participation. I speak for President Norton and Dr. Rosser and their associates in expressing our appreciation to the University and the University community for their very great contribution in working out these collaborative arrangements.

IDA has another reason for being happy with the choice of Princeton's hospitality: I refer to the exceptional resources in mathematics, perhaps unexcelled in the world, which are concentrated here in the University and in the Princeton community. Altogether we feel the auspices to be particularly benign for our undertakings here.

Woodrow Wilson once observed that "Government is not a machine, but a living thing. It falls, not under the theory of the universe, but under the organic life. It is accountable to Darwin, not to Newton." The Institute for Defense Analyses represents a small organic extension of Government—a mutation which provides new opportunities for our Government to draw upon the resources of the nation's academic and scientific communities in a way accountable both to Newton and to Darwin and that is beneficial to these communities as well as to the Federal Government. Since the war we have seen many ingenious methods devised in the area of what Don Price has called "federalism by contract" which have served to make available to national policy making and the national Government

## VON NEUMANN HALL DEDICATION UNCLASSIFIED

advisory services otherwise not easily accessible. This enterprise here in Princeton represents one of these novel and, we venture to hope, mutually beneficial arrangements which brings a new element of strength to our national life.

President Goheen, we are happy and privileged to be in this university community and to have the cooperation of your institution and your colleagues.

*Dr. Robert F. Goheen . . .*

Our gathering this morning, widespread though we are in the institutions we represent and diverse perhaps in our interests, brings us together to pay honor in common to the memory and influence of a rare individual.

Three decades ago John von Neumann, at the age of 26, accepted Princeton University's invitation to join its faculty as professor of mathematical physics. A very few years later (1933), when the Institute of Advanced Study was brought into being, he was appointed one of its founding professors. Thereafter, whether here in Princeton or commuting to Los Alamos or Santa Monica, or serving with scientific commissions, or temporarily making his home in Washington, Professor von Neumann considered himself a Princetonian and loved the ways of life that are the hallmarks of this academic community.

All of us here are deeply aware of Professor von Neumann's scientific legacy, of his salient contributions to man's knowledge, of his devotion to the principles by which free men live. We who are Princetonians recall with especial pride the honors that rightfully came to him: The medal for Merit in 1947; the Medal of Freedom in 1956; and the Enrico Fermi Award,—the citation attending the last rightly stressed that he more than anyone else foresaw the important and necessary role which high-speed computing machines would play, not only in the control and use of atomic energy but also in the general advancement of the sciences.

The vital influence he exerted in all that he did was expressed wonderfully well in the dismal February of 1957 by one of his close friends and associates who is with us today, Dr. Robert Oppenheimer:

"To his many friends, his students, his colleagues, John von Neumann was the highest and liveliest intelligence they were ever to encounter. A mathematician of immense scope and power, he contributed to many fields of learning and *created others*. He was a masterful abstract analyst, with an unparalleled sense for practical invention, so that he enriched learning and practice equally. His sober and often melancholy realism was tempered by great warmth and generosity. We know no one like him."



## UNCLASSIFIED VON NEUMANN HALL DEDICATION

Recollections of this outstanding man of science, who could recite *Faust* from memory and whose outreach was immense, have become a kind of living legend in our community. For myself, I especially enjoy an anecdote which I first got from a colleague but later saw repeated in a published tribute.

One evening at the von Neumanns' Westcott Road home, so the story goes, Professor von Neumann and an eminent Byzantinist were discussing a little-explored corner of history and came to argument over a date. The historian said it was this; von Neumann insisted it was that. Eventually, as scholars should, they looked it up, and von Neumann was right. Some time later, when again invited to the von Neumann house, the historian is said to have said: "I'll come if Johnny promises not to discuss Byzantine history. Everybody thinks I am the world's greatest expert in it, and I want them to keep on thinking so".

It was fifteen years ago that this warm, many-sided individual, anticipated the creation of an organization to serve purposes such as The Institute for Defense Analyses now serves. In a memorandum dated September 5, 1945—in which he was analyzing the one high-precision electronic machine then approaching completion—he wrote:

"There are many important problems in hydrodynamics, aerodynamics, celestial mechanics, and in various other fields, which are practically inaccessible to the present methods of abstract mathematical analysis, and for which the capacity of human computing machines, or of existing, non-electronic computing machines, is absolutely inadequate. These problems can only be dealt with by machines which possess intrinsic speeds that can only be reached by electronic procedures. Such speeds render any intelligent human intervention, while the machine is working, impossible, and therefore they necessitate a complete automation of the device."

He went on to suggest that the construction and operation of an "all-purpose machine" should be undertaken immediately by a purely scientific organization, as opposed to existing governmental or industrial agencies. He emphasized the desirability of planning "without any inhibitions," for free operation governed principally by scientific considerations.

So also, writing three days after the Japanese surrender, Professor von Neumann urged that "the Government needs the help of a scientific institution" for these purposes, and he foresaw the creation of some future, central postwar research agency that might well be economically self-supporting. Again he cautioned that its independence and ability to exert a directing influence on future developments were matters of vital concern.

This morning—an even decade and a half after Professor von Neumann, peering into the future, suggested this kind of charter for

## VON NEUMANN HALL DEDICATION UNCLASSIFIED

what is now the Institute for Defense Analyses—we sit, as it were, before the translation into reality of some significant part of his vision. And I find myself greatly moved by the devotion and respect which are apparent in this gathering—devotion and respect for the man whose memory we honor, devotion and respect for those qualities of mind and character which he displayed in his all too short lifetime, and of which the Institute for Defense Analyses, with its demonstrated achievements and its rich promise for the future, is such an appropriate expression.

UNCLASSIFIED

## Pattern Recognition

BY WALTER W. JACOBS

*Unclassified*

*The broad class of pattern recognition problems is considered, and the example of handwritten signatures is used to elucidate the general problem. A model is presented which shows the relation of pattern recognition to communications, and the structure of recognition procedures is discussed.*

Human sensory perception seems to involve pattern recognition in a fundamental way. When we examine the physiological processes involved in the perception of shapes and colors, sounds, textures, and so on, we find that in these processes there are large numbers of receptor elements affected by any stimulus. Starting with the arrangement of the affected elements, the brain arrives at an appropriate image, although we have little knowledge of how this is accomplished.

*Webster's New International Dictionary* defines pattern as "... an arrangement of parts, elements or details that suggests a design or orderly distribution." A definition that is closer to our requirements is: a design or orderly structure that underlies an arrangement of parts, elements or details. To recognize a pattern is to detect or identify the structure associated with the particular arrangement or occurrence.

The term "pattern recognition" is sometimes restricted to refer to the identification of shapes. We are using it in a broader sense, to include, for example, speech recognition and even such other problems as identifying a piece of music, a poem, a face or a voice.

In speeding up the processing of information, the mechanical recognition of patterns is becoming increasingly necessary. Various forms of this problem are being worked on. In order to provide a firmer foundation for such work, a model of the general recognition problem is presented here; this model is explicit enough to provide a formulation for mechanical recognition, and at the same time it appears broad enough to encompass human recognition.

The principal aim of such a model is to indicate the conditions that should be satisfied if a successful recognition procedure is to be achieved. Thus it provides a basis for evaluating partial attacks on the problem.

UNCLASSIFIED

UNCLASSIFIED

PATTERN RECOGNITION

## THE PROBLEM OF SIGNATURE RECOGNITION

The exposition will be organized around a specific example—the mechanical recognition of a handwritten name. The example has been chosen because of its concrete nature, and because it is extremely familiar to everyone.

In recognizing a signature, we may be trying to answer one of a number of possible questions. What is the name of the signer? Does the signature correspond to the standard for a checking account in a particular bank? Is it a valid or a forged signature? Each of these questions gives rise to a different recognition problem, and it should be clear that the corresponding procedures need have little or nothing in common. It is necessary to be very explicit about the problem to be solved in order to avoid tackling too much or accomplishing too little.

The discussion will deal with the first of these questions: identifying the name. It is assumed that the only information available for the problem consists of two lists, one containing 250 first names, and the other 4,000 surnames. (We suppose, to eliminate a complication that would add nothing to the exposition, that there is no middle initial.) No standard signature, such as would exist in the bank problem, is provided.

The objective in discussing this problem is to illustrate and illuminate the general situation. For present purposes, it is of little concern whether this form of the problem is of practical interest, or whether the approach to be described is feasible.

## THE COMMUNICATION MODEL OF PATTERN RECOGNITION

Our model is based on an analogy between the usual communication situation and the pattern recognition problem. In communication, we deal with messages, sent and received. The original *message* is converted to a *signal* or other physical form and is transmitted along some communication channel. The channel is in general “noisy”, and the signal is distorted or modified. It is then received and recorded, further degradation of the signal occurring in the process, and the resulting *record* is used to obtain information about the original message. By interpreting “message”, “signal” and “record” in a somewhat more general sense, we can identify these same elements in pattern recognition.

In communications, the origination and transmission of the message are usually intentional; in pattern recognition, however, this is too restrictive. The criminal who leaves his fingerprints at the scene of his crime is unwittingly sending a message to the detective. To the latter, the identity of the criminal corresponds to the pattern underlying the fingerprints.

UNCLASSIFIED

W. W. JACOBS

UNCLASSIFIED

Similarly, in the signature problem, the individual writing his name is originating a message. The light reflected by the signature plays the role of the signal, and the receiving element may be the retina of the eye or the photosensitive component of a mechanical scanning device. In the form of the problem being discussed, the name is the desired pattern.

These two examples illustrate what is meant by "pattern" in general. A particular pattern, such as the name "John Smith", can be represented by many different messages; not only can the signer vary the size and form of his signature, but also there may be different people writing the name. However, in the situation being discussed, all possible signatures fall into  $(250)(4,000) + 1 = 1,000,001$  different classes, corresponding to the possible pairs of names on the given lists or to the additional case—the "null pattern"—when one or both names are not listed (or perhaps what is being examined is not even a signature).

Because handwriting is often bad, and because "noise" further obscures what is written, no recognition procedure can be uniformly successful in assigning a record to its pattern. It takes only a smudge to make "Jean" practically indistinguishable from "Joan", or "South" from "Smith". What the procedure can do is divide up the set of possible records into classes or categories, each class corresponding to a single pattern. For example, one such class would contain every record which is assigned to "John Smith". This dividing up should be done so as to minimize the effect of incorrect recognition.<sup>1</sup>

Some writers have used this dividing up of the set of records into classes as the basis for a definition of pattern. It becomes "that property which all the records of a single class have in common." The definition is unsatisfactory, and involves a confusion between the recognition procedure, on the one hand, and the success with which a given mechanism achieves the intended assignment of records to classes, on the other. This becomes clearer when it is realized that the definition excludes any notion of *pattern structure*, a term which encompasses all the knowledge about the patterns which is not present in the totality of records. As the example will show, it is this external knowledge on which a recognition procedure is based.

#### MESSAGE, SIGNAL AND RECORD

To look at a communication situation and to specify the place where the message is in existence and entering the transmission proc-

<sup>1</sup> Procedures which allow for more categories of records than there are patterns may be required if, under appropriate conditions, an indecisive outcome is desirable. This involves additional considerations, and is not discussed further, although the present treatment can readily be modified to handle it.

UNCLASSIFIED

PATTERN RECOGNITION

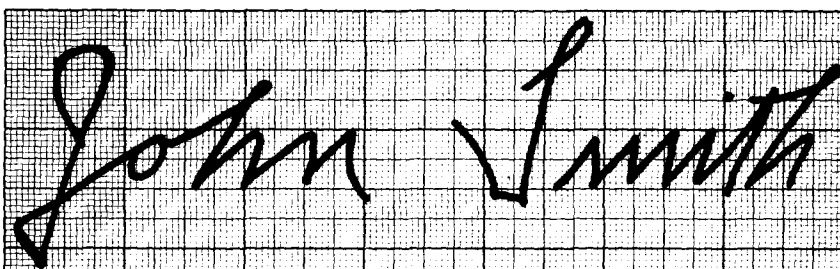
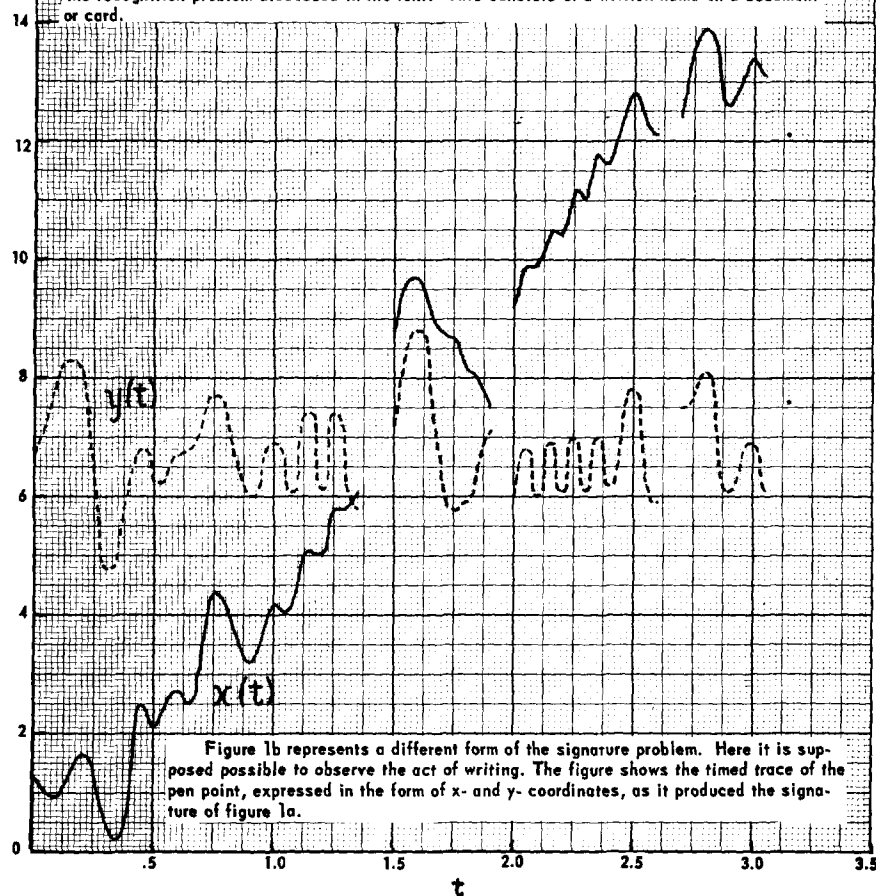


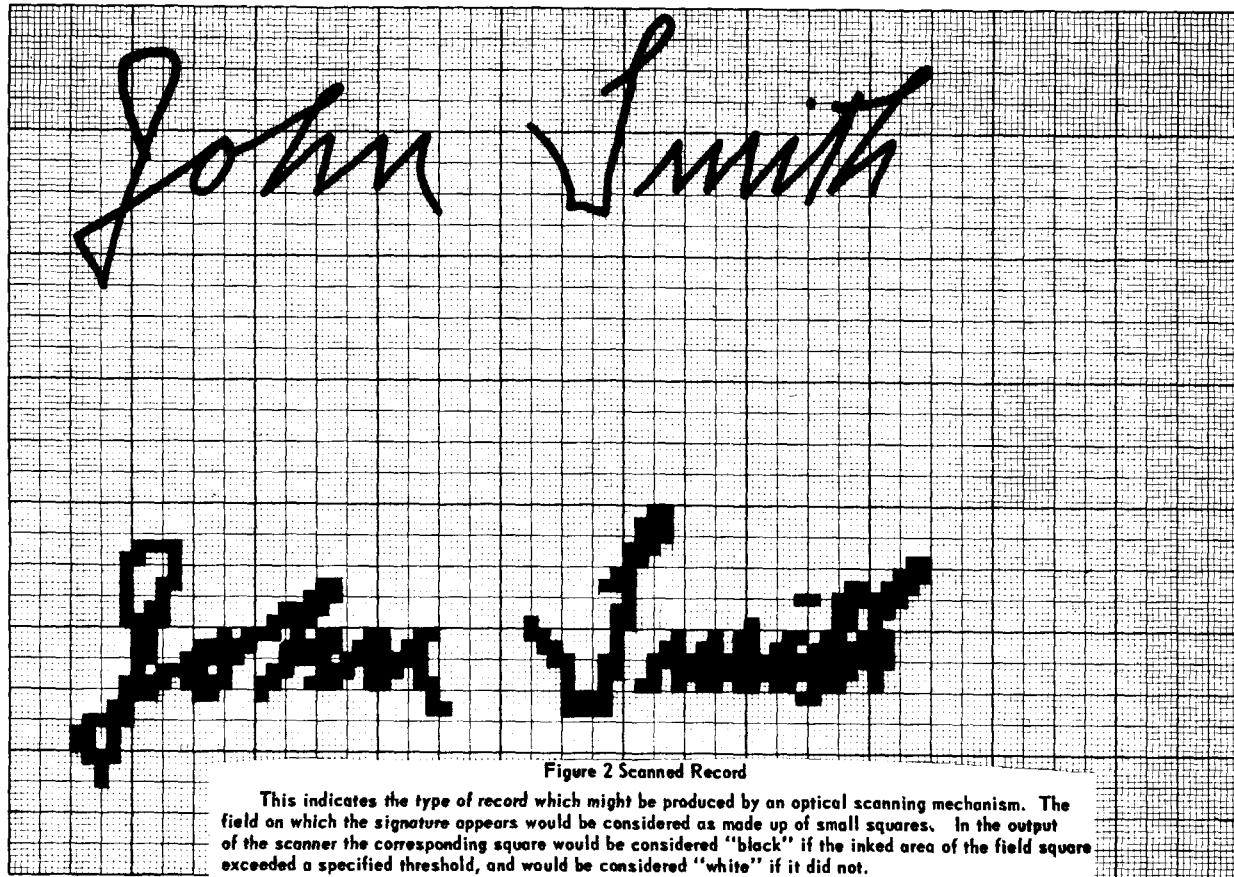
Figure 1a shows the signature "John Smith" in the form which is assumed available for the recognition problem discussed in the text. This consists of a written name on a document or card.



UNCLASSIFIED

W. W. JACOBS

UNCLASSIFIED



UNCLASSIFIED

PATTERN RECOGNITION

ess, often calls for an arbitrary choice. One may say that transmission commences with the conversion of a written message to electrical form; it is just as reasonable to say that it begins with the conversion of a mental message to verbal or written form.

The same kind of choice arises in recognition situations, and one can be somewhat free in specifying the point at which the message exists and the dynamic process which constitutes the signal has begun. The signal intervenes physically between the originator and the recipient, and the form of signal available to the latter depends on his relative location and other conditions. In the class of recognition problems represented by a given procedure, it is useful to think of reception conditions as relatively fixed.

In the example of the handwritten name, the usual physical effect will be a piece of paper or cardboard containing the writing. This may be called the "static" signature (Fig. 1a), and this is the kind of signal which will be considered in the discussion here. If the conditions of the problem permitted us to observe the act of signing, then the signal would more properly be regarded as the timed succession of pen positions, as indicated in Fig. 1b. The two problems are evidently different, although the same pattern is to be recognized. A recognition procedure based on the second type of signal could make use of knowledge (about the succession and direction of strokes, for example) which is not directly available with the static signature.<sup>2</sup>

Before a recognition procedure can be applied to the signal, it must first be received and transcribed. The resulting record is the input to the procedure, and summarizes all of the information that applies to the particular occurrence of the problem. Information about the pattern structure—such knowledge as the fact that names are produced by continuously-drawn strokes, and that they consist of letters in pronounceable combinations—is not present in the record, but is collateral to it.

Figure 2 shows a record of the signature such as might result from scanning the signal with a facsimile device. The form of the record is chosen by the recognizer, within the limitations imposed on him by his access to the signal and the technical devices available to him. The signal itself is a physical phenomenon which he cannot control.

<sup>2</sup> After the present paper was written, the author was shown an unpublished manuscript, "Machine Reading of Handwriting", by L. S. Frishkopf and L. D. Harmon of Bell Telephone Laboratories, which discussed a proposed machine procedure for recognizing handwritten material from the time plot of its  $x$  and  $y$  coordinates. The approach is an excellent illustration of the general class of procedures described here.

UNCLASSIFIED



W. W. JACOBS

UNCLASSIFIED

## VARIATION, NOISE AND PROBABILITY CONSIDERATIONS

An individual will produce many versions of his signature, reflecting such sources of variation as space available, muscular control, and so on. With upwards of 100,000,000 Americans who can write, producing dozens of possible signatures each, the number of different messages in our problem is in the billions. These are grouped, by the task set, into 1,000,001 patterns. The differences among the messages in a pattern class constitute the *variation*.

When John Smith undertakes to produce an instance of his signature, a particular message in the pattern corresponding to his name is originated. However, the actual signal is not uniquely determined by that message because of *noise*. The table on which he is writing may shake, the ink may blot or smudge, and dirt or moisture may further alter the written form of the signature before it reaches the recognition process.

The act of producing the record itself will introduce more noise to obscure the information present. The effect which results from the discrete field of the scanner is graphically shown in Fig. 2; however, noise is inevitable in any device, whether it operates discretely or continuously.

Because of the intervention of variation and noise, it cannot be assumed that every record can be unambiguously assigned to one and only one pattern. As has already been pointed out, we frequently encounter cases where even the human recognition procedure fails to obtain a decisive answer. Our model assumes that every pattern gives rise to a definite probability distribution over the set of possible records.<sup>3</sup>

The recognition procedure must take account of the probability distributions for the various patterns. Often these will not be given a priori, but must be estimated on the basis of samples of records for which the corresponding patterns are known.

The general recognition problem, in terms of the model which has been presented, is therefore seen to fall into the well-known category of statistical problems in which we have a single sample drawn from one of a finite set of populations, and wish to "estimate" the population from which it originated. There are, therefore, two distinct aspects to any recognition procedure: one concerned with the statistical decision that must be made, and the other involved with the means of transcribing the original signal and of physically carrying out on the resulting record the statistical calculations that are necessary.

<sup>3</sup> As discussed in the Appendix, this assumption appears to be necessary in order that the recognition problem be well-defined.

UNCLASSIFIED

PATTERN RECOGNITION

## THE STRUCTURE OF THE RECOGNITION PROCEDURE

When there are many patterns, or when there is a considerable amount of variation within the individual patterns—in other words, whenever the number of possible messages is very large—the statistical rules that would assign each record to its appropriate pattern are too complex to carry out in a single step. In this case, the recognition procedure is more appropriately considered as a series of operations.

There are four well-marked functions or aspects in recognition, and we call them Representation, Extraction, Classification, and Integration. They are indicated in the flow-chart of Fig. 3. These aspects appear to be necessary in any non-trivial problem, and one can use them to appraise the extent of progress that is made in any practical proposal for a mechanical recognition device.

If the published material on character recognition and other problems is evaluated on this basis, in most cases it appears that only a part (and often the easier part) of the recognition procedure has actually been attacked. Only where variation and noise can be rigidly controlled—as in examples of character recognition where the method of printing is precisely specified—has much headway been made, and even in these relatively simpler problems the procedures described appear to contain some serious gaps.

The four aspects are discussed in turn in what follows. The logical flow-chart of Fig. 3 is not intended to indicate a corresponding physical separation of function in an actual device; it is quite possible that in a particular procedure a single mechanism could effectively combine two or more functions.

## REPRESENTATION

The determination of the method of recording the signal constitutes the first aspect of the recognition procedure. This step is called *representation*, and it may also be thought of as selecting the form of the record.

This same first stage exists in human recognition. The perception of a shape, for example, commences with the stimulation of certain of the discrete array of receptor cells in the retina of the eye.

We have already seen that representation introduces noise. The signal-to-noise ratio may be increased within limits by increasing the faithfulness of the recorder. In Fig. 2, if the field being scanned were divided into a larger number of cells, the record would more closely approximate the signature.

It is possible to incorporate a noise "filter" or "suppressor" into a recording device, but this merely combines with the representation stage a function that properly should be considered part of a later

UNCLASSIFIED

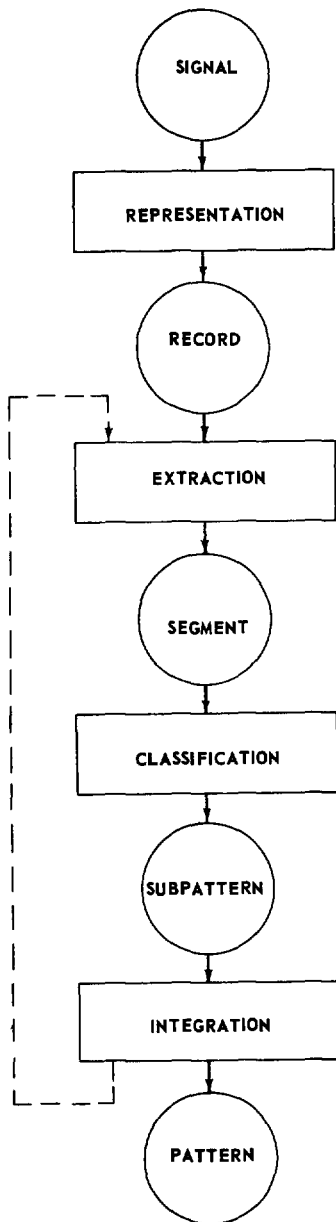


Figure 3  
Structure of the General Recognition  
Procedure

This flow-chart indicates the functional nature of the recognition procedure rather than the organization of physical components of a recognition device. The rectangles represent aspects or functions of the procedure, the circles inputs and outputs. The dashed line suggests the iterative or sequential approach which is described.

UNCLASSIFIED

PATTERN RECOGNITION

section of the recognition procedure. Paradoxically, the recognition procedure can sometimes be simplified by using an extremely coarse scan, thereby passing up much of the information available in the effect. Here again the representation stage is being combined with later parts of the recognition procedure.

It is useful to discuss the informational aspect of the problem somewhat further. Because 1,000,001 is approximately  $2^{20}$ , a signature which has been correctly assigned to its pattern—i. e., has had its name identified—has contributed about 20 bits of information relevant to the name identification. It contains far more information than this, some of which would be relevant to other problems. The use of pattern structure and its implied redundancy makes it possible to discard much of the additional information. However, the function of discarding information should be kept conceptually distinct from the function of representing it, even though a single physical device may simultaneously perform both functions.

#### EXTRACTION

The remainder of the recognition procedure operates on the record as input, and yields an estimate of the underlying pattern. While it is theoretically valid to consider this estimate as a statistic calculated from the record by a single mathematical operation, in practice it is often important to break down the process.

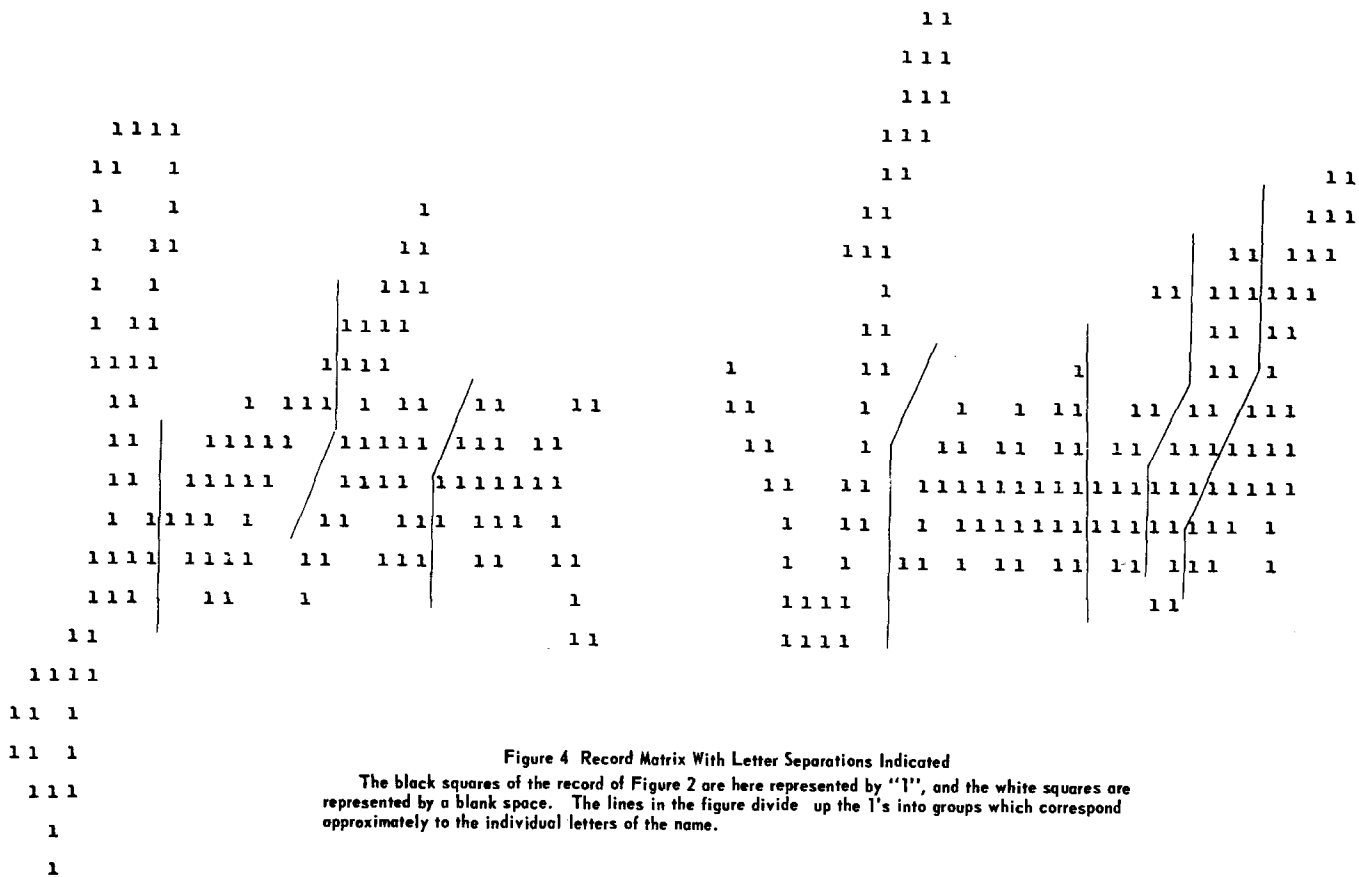
The estimation of the pattern from the record begins with a function that may be called *extraction*. We will first consider what this involves, and then indicate some of the difficulties associated with carrying it out. Extraction consists of selecting a part of the record, or *segment*, in such a way as to reduce or eliminate the effect of a source of variation in the original set of messages. There are two distinct ways in which this can take place. The segment may correspond to a group of patterns so chosen that some variation is cancelled out; such a group is called a *subpattern*. For example, if all patterns are grouped according to the first name, and the segment includes only the corresponding part of the record, then the subpatterns are the first names. In this partial problem there is much less variation than in the full problem involving a single complete pattern. The reduction is much greater if the segment includes only that part of the record corresponding to the initial letter.

The second type of extraction aims at counteracting variation within the pattern classes by selecting parts of the record that correspond to some standard or canonical feature of the entire class. Adjustments for variations in size, registration or orientation of the record are of this type.

UNCLASSIFIED

W. W. JACOBS

UNCLASSIFIED



**Figure 4 Record Matrix With Letter Separations Indicated**

The black squares of the record of Figure 2 are here represented by "1", and the white squares are represented by a blank space. The lines in the figure divide up the 1's into groups which correspond approximately to the individual letters of the name.

UNCLASSIFIED

PATTERN RECOGNITION

The term "segmentation" is frequently used for the first type of extraction. However, it seems desirable to replace this term, both because it does not fit the second type, and because it seems to imply that the entire record is divided up into segments at one stroke; this latter restriction is unnecessary.

Extraction is often a troublesome stage, especially if it is done in the form of complete segmentation. Some of the reasons for this are exemplified in Fig. 4, which shows the record of Fig. 2 (expressed as a matrix of 0's and 1's with the 0's suppressed for clarity).

In the figure, the approximate segmentation into letters is indicated by lines.<sup>4</sup> But the matrix, although set out in a rectangle, will as a practical matter be read linearly, and most probably either by rows or by columns. Thus, the segments will not in general be connected pieces of this linearly described record. Either the extraction rule must be a complicated one, or the segments will not correspond perfectly to single letters.

There are other problems in extraction besides non-linearity. The segments are not usually independent, because the way a letter is written will depend on what its neighbors are. Also, there is frequently an intrinsic ambiguity characteristic of handwriting, which is illustrated in deciding where to terminate the second letter of the second name: is this an i or r, n or u, or m?

The kinds of difficulty pointed out here can arise in many recognition problems. In trying to avoid these difficulties, the attempt is sometimes made to dispense with extraction and work with the record as a whole. However, except in rather trivial cases, at least the second type of extraction is still necessary—and because this may involve a less specific kind of pattern structure, it may be no easier to handle properly.

Consider, for example, a variation of the signature problem in which a file of standard signatures is available for comparison. One can conceive of an approach which matches the record as a whole against each standard. In order to do this successfully, the matching process must relate corresponding parts of the record and the standard. But parts of a signature vary in their relationship to each other; in comparing signatures by eye, we find it necessary to adjust for differences in spacing or size which interfere with any simple basis for comparison. Thus, extraction is still required to make the parts of the record correspond to those of the standard.

<sup>4</sup> As an instance of the puzzles of human perception, the reader should note how much more easily the name may be recognized from the record as shown in Fig. 2 than from the informationally equivalent form in Fig. 4.

UNCLASSIFIED

W. W. JACOBS

UNCLASSIFIED

In the experimental stage of developing such a procedure, a correlation technique relying on the presence of "above-the-line" letters (b, d, f, h, k, l, t) and "below-the-line" letters (f, g, j, p, q, y, z) might be tried. The extraction might involve location and scale adjustments of the record so as to produce coincidence at both the beginning points and ending points of the observed and standard signatures. However, writing does not always space its letters uniformly, and we can therefore conclude that the indicated approach would have a higher probability of error than could be attained with a more sophisticated extraction. Whether the errors could be tolerated in order to keep the procedure simple would depend on the specific circumstances of the actual application.<sup>5</sup>

Let us return to the signature problem which we have been using as an example, and use it to emphasize an important point about subpatterns. These are not predetermined by the structure of the patterns, but are selected in the course of developing an efficient recognition procedure. In the signature problem names are composed of letters, and letters of strokes, but this fact does not compel us to use letters or strokes as our subpatterns.

A relatively simple way to start the process would be to extract as a segment the part of the record that corresponds to the initial letter of the first name. However, the underlying subpattern would be not the letter itself, but a properly chosen group of letters. One such group might be the letters which as capitals have a loop below the line: J, Y, Z.<sup>6</sup>

In summary, extraction operates on the record and produces a segment, in such a way as to offset a substantial part of the variation present among the original set of possible messages. It may do this by standardizing, so that the resulting segment still is used to estimate the entire pattern. Alternatively, the segment may relate to a subpattern of the original pattern.

#### CLASSIFICATION

The determination of the estimated subpattern from a segment of the record is called *classification*; it constitutes the third aspect of the recognition procedure. It differs from the problem of estimating the entire pattern from the record in only one respect: that no further extraction takes place. At this stage a manageable piece of the

<sup>5</sup> A typical "practical" solution might be to weight the errors in favor of rejection, and use human inspection of the rejects. Since the back-up inspection is part of the "mechanical" recognition procedure, this solution might turn out to be much more expensive to operate (although much cheaper and quicker to develop!) than a more sophisticated machine.

<sup>6</sup> How to continue from this start is discussed below under integration.

## UNCLASSIFIED PATTERN RECOGNITION

original problem has been cleaned up. The input may be thought of as an observed vector (or a continuous counterpart); the result of the classification is an estimate of the subpattern.<sup>7</sup>

In order to determine how a machine is to carry out the classification aspect of a recognition procedure, two interrelated tasks must be accomplished. The first task is to decide which subpattern each possible segment should be assigned to; and the second is to produce a mechanism or calculation which actually accomplishes or sufficiently approximates the desired assignment. The principal consideration in handling the first task is the risk associated with erroneous assignments; in the second task, the practicality of the assignment procedure.

We referred earlier to the literature on pattern recognition problems. The partial attacks we mentioned are largely concerned with classification, and within this area primarily with the carrying out of a specified assignment function.<sup>8</sup> Relatively few writers give any recognition at all to the task of choosing the assignment function. In some cases this seems to stem from the view that it is someone else's job to eliminate (in our terminology) all of the noise that exists in the channels of communication, for in the absence of noise the correct assignment is presumably known.

The determination of a suitable assignment function is analogous to the standard statistical problem of developing a test of hypotheses. The segments of the records are analogous to vector samples, and the subpatterns correspond to the statistical universes from which the samples are drawn. In our example, the subpatterns for the initial segment are groups of capital letters, such as the group (J, Y, Z); and an additional one, the "null" subpattern, corresponding to the case where no underlying letter is present, for example, because the extraction was incorrectly done.

The methods of deriving suitable statistics for the classification problem belong to decision theory. In order for these methods to be applied, however, two assumptions must hold. These assumptions are: first, that each subpattern gives rise to a valid probability dis-

<sup>7</sup> It can be questioned whether this oversimplifies the situation; might not the proper output be a set of probabilities or other scores assigned to the various subpatterns? Such a modification is not necessary, but to explain the justification for this assertion, a long and difficult philosophical and mathematical digression would be required. In any case, those who prefer may interpret "estimated subpattern" as a vector of posterior probabilities; nothing essential in the remaining discussion is affected.

<sup>8</sup> The Perceptron research, and related investigations, treat the problem of *learning* an assignment function from a sample of assignments. The pre-existence of the function, in implicit form at least, is apparently assumed.



tribution of the corresponding segments; second, that either the knowledge of pattern structure yields complete specifications of these distributions or there are available adequate samples to estimate any unknown parameters.

The importance of these assumptions is underlined by the fact that many attempts to develop recognition procedures are carried out under laboratory conditions, with noise kept to a minimum. The result is that no information is obtained about the probability distributions that hold under actual conditions, and the procedures that apply in the laboratory may be of little use outside.

It is sometimes claimed that the question of a proper statistic for classification is not important: that there exists some transformation of the segment which will expose an invariant, easily identified characteristic of the subpattern. This assumption has generally proved too optimistic even in such relatively favorable situations as the recognition of characters from a fixed font. In the case of written letters, even the human will make some proportion of bad identifications without the help of context to correct him. In our handwriting example, it is evident from Fig. 4 that once the letter "n" has been isolated, there can be no test which clearly differentiates it from the letter "u". However, as soon as it is admitted that no perfect test can exist, it becomes important to specify one that keeps down the probability of error.

It should not be assumed, because we have said that methods exist for the determination of classification statistics when the proper assumptions are satisfied, that the task is always a trivial one. Even when the number of subpatterns is as small as three, the theoretical difficulties can be severe and the practical ones are worse, unless the subpatterns have been well chosen. As the number of subpatterns becomes large, the difficulties become overwhelming. This may help to account for the almost universal hope in such cases that some panacea will, by great good luck, do an adequate job.

Trying to avoid extraction by working with the record as a whole, one has to find a method of classification with large numbers of underlying patterns. All of our experience supports the conclusion that most of the time it is better to break a problem into smaller parts than to struggle with the undivided problem.

With our handwriting example, we shall prudently avoid any discussion of possible classification statistics. For the purposes of the rest of the discussion, we assume that such a statistic has been applied, and has produced an estimated subpattern for the initial segment.

UNCLASSIFIED

PATTERN RECOGNITION

## INTEGRATION

In the usual recognition problem a series of extraction and classification steps will give rise to a sequence of subpatterns. Making the sequence of subpatterns "add up to" an answer to the original problem is the job of the *integration* function in the recognition procedure. Integration must therefore control the sequencing of the extraction and classification, and handle any feedback that is involved in the recognition.

Consider the handwriting example, and suppose that the initial segment has been classified as belonging to the subpattern of "tailed" capital letters, i. e., is one of J, Y, Z. This produces, out of the list of 250 first names, a list of perhaps 20 possibilities for the remainder of the first name: ames, ohn, oan, vonne, ves, achary, elda, and so on. Clearly the use of such a small and specific list of possibilities can bring into consideration a new group of practical possibilities for the extraction and classification of the second segment. As a consequence of this type of feedback, every iteration should be able to restrict itself to a small number of subpatterns.

If the extraction of segments and their classification proceed independently rather than iteratively, and if there is no effort to correct errors in these stages either by context or by error detection and iteration, then integration is a relatively trivial step. For problems of any complexity, where errors in the earlier aspects cannot be allowed to cause the procedure to fail, integration is a major aspect of the procedure.

If integration is to control iterative processes and use earlier results to make decisions about later ones, then error detection and correction become possible. For example, the set of letters J-o-h-u is readily recognized as "John". Whether J-o-h-f should be treated as a name not on the list depends on probability distributions of that outcome under the two subpatterns "John" and "not listed", as well as on the consequences of the two possible way of making an erroneous decision.

## CONCLUSION

We have described a model of pattern recognition, based on treating the problem as related to communication theory. This model leads to a structure for recognition procedures in general, and provides a basis for evaluating the thoroughness with which a proposed procedure attacks the various aspects of the recognition task.

Although the discussion may have appeared to stress the difficulties of such functions as extraction and classification, our purpose has been primarily to warn against a tendency to gloss over or wish

UNCLASSIFIED

W. W. JACOBS

UNCLASSIFIED

away certain aspects of the job. We feel that these difficulties are surmountable, once they are squarely faced.

In fact, when the problems are formulated concretely enough, even such currently unmanageable tasks as identifying handwritten material by machine can begin to look quite feasible. The use of iterative procedures, which at all points deal with a relatively small set of subpatterns, and apply these to narrow the problem successively, seems to represent the most hopeful direction for continued exploration.

## APPENDIX

It appears useful to restate in mathematical language the model that has been presented in the preceding pages. This model involves

$M$  = the space of *messages*  $m$ .

$S$  = the space of *signals*  $s$ .

$R$  = the space of *records*  $r$ .

Associated with each point of  $M$  is a conditional probability measure on  $S$ :

$$Pr[s|m]$$

and with each point of  $S$  a conditional probability measure on  $R$ :

$$Pr[r|s].$$

These induce a conditional probability measure

$$Pr[r|m] = \int_S Pr[r|s] d Pr[s|m].$$

A *pattern*  $P$  is a partition of  $M$ : that is, a set  $P_i$  of non-overlapping classes of messages which together exhaust  $M$ .

$$M = P_0 + P_1 + \dots + P_k, \quad P_i P_j = \emptyset.$$

A decision procedure  $D$  for recognition is a corresponding partition of  $R$ : that is a set  $D_i$  such that

$$R = D_0 + D_1 + \dots + D_k, \quad D_i D_j = \emptyset.$$

In order that the validity of a decision procedure can have any meaning, the probability measures

$$Pr[r|P_i]$$

UNCLASSIFIED

PATTERN RECOGNITION

must exist. This requires that there be a set of relative probability measures over each of the subspaces  $P_i$ ; if these measures are denoted by  $\mu_i$ , then

$$Pr[r | P_i] = \int_{P_i} Pr[r | m] d\mu_i(m).$$

However, the partition  $P$  is specified by the recognizer, and therefore the relative measures must exist for any partition  $P$ . This can happen only if there is a measure  $\mu$  defined on  $M$ , with

$$\mu_i(m) = \mu(m) \div \mu(P_i) \text{ whenever } \mu(P_i) > 0.$$

The problem now reduces to the typical estimation problem of decision theory, with the classes  $P_i$  corresponding to the states of nature or hypotheses, the sets  $D_i$  the actions or estimates, and the  $\mu(P_i)$  the prior probabilities.

UNCLASSIFIED

~~CONFIDENTIAL~~

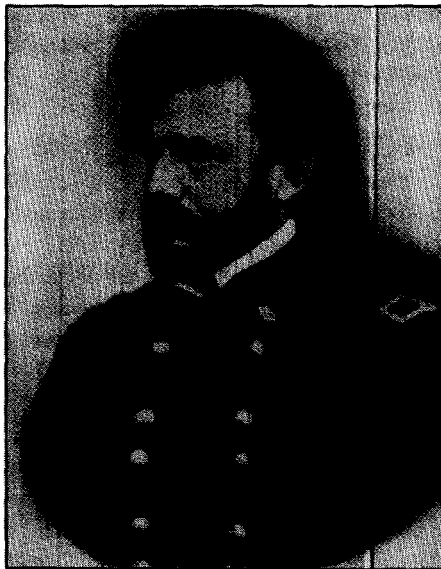
## Introduction to Cryptology—IV

BY WILLIAM F. FRIEDMAN

*Confidential*

*Cryptology in the Civil War*

A detailed account of the codes and ciphers of the Civil War in the United States of America can hardly be told without beginning with a bit of biography about the man who became the first signal officer in history and the first Chief Signal Officer of the United States Army, Albert J. Myer, the man in whose memory that lovely little U. S.



BRIGADIER GENERAL ALBERT J. MYER

Army post adjacent to Arlington Cemetery was named. Myer was born on 20 September 1827, and after an apprenticeship in the then quite new science of electric telegraphy he entered Hobart College, Geneva, New York, from which he was graduated in 1847. From early youth he had exhibited a predilection for artistic and scientific studies, and upon leaving Hobart he entered Buffalo Medical College, receiving the M.D. degree four years later. His graduation thesis, "A Sign Language for Deaf Mutes," contained the germ of the idea he was to develop several years later, when, in 1854, he was commissioned a 1st Lieutenant in the Regular Army, made an Assistant

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

Surgeon, and ordered to New Mexico for duty. He had plenty of time at this far-away outpost to think about developing an efficient system of military "aerial telegraphy," which was what visual signaling was then called. I emphasize the word "system" because, strange to say, although instances of the use of lights and other visual signals can be found throughout the history of warfare, and their use between ships at sea had been practiced by mariners for centuries, yet down to the middle of the 19th Century surprisingly little progress had been made in developing methods and instruments for the *systematic* exchange of military information and instructions by means of signals of any kind. Morse's practical system of electric telegraphy, developed in the years 1832-35, served to focus attention within the military upon systems and methods of inter-communication by means of both visual and electrical signals. In the years immediately preceding the Civil War, the U. S. Army took steps to introduce and to develop a system of visual signaling for general use in the field. It was Assistant Surgeon Myer who furnished the initiative in this matter.

In 1856, two years after he was commissioned assistant surgeon, Myer drafted a memorandum on a new system of visual signaling and obtained a patent on it. Two years later, a board was appointed by the War Department to study Myer's system. It is interesting to note that one of the officers who served as an assistant to Myer in demonstrating his system before the board was a Lieutenant E. P. Alexander, Corps of Engineers. We shall hear more about him presently, but at the moment I will say that on the outbreak of war, Alexander organized the Confederate Signal Corps. After some successful demonstrations by Myer and his assistants, the War Department fostered a bill in Congress, which gave its approval to his ideas. But what is more to the point, Congress appropriated an initial amount of \$2,000 to enable the Army and the War Department to develop the system. The money, as stated in the Act was to be used "for manufacture or purchase of apparatus and equipment for field signaling." The act also contained another important provision: it authorized the appointment, on the Army staff, of one Signal Officer with the rank, pay, and allowances of a major of cavalry. On 2 July 1860, "Assistant Surgeon Albert J. Myer (was appointed) to be Signal Officer, with the rank of Major, 27 June 1860, to fill an original vacancy," and two weeks later Major Myer was ordered to report to the Commanding General of the Department of New Mexico for signaling duty. The War Department also directed that two officers be detailed as his assistants. During a several months' campaign against hostile Navajos, an extensive test of Myer's new system, using both flags and torches, was conducted with much success. In

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

October 1860, a Lieutenant J. E. B. Stuart, later to become famous as a Confederate cavalry leader, tendered his services to aid in signal instruction.

Less than a year after Major Myer was appointed as the first and, at that time, the only Signal Officer of the U.S. Army, Fort Sumter was attacked and, after a 36-hour bombardment, surrendered. The bloody four-year war between the North and the South began. The date was 14 April 1861. Myer's system of aerial telegraphy was soon to undergo its real baptism under fire, rather than by fire. But with the outbreak of war, another new system of military signal communication, signaling by the electric telegraph, began to undergo its first thorough test in combat operations. This in itself is very important in the history of cryptology. But far more significant in that history is a fact that I mentioned at the close of the last lecture, *viz*, that for the first time in the conduct of organized warfare, *rapid and secret military communications on a large scale became practicable*, because cryptology and electric telegraphy were now to be joined in a lasting wedlock. For when the war began, the electric telegraph had been in use for less than a quarter of a century. Although the first use of electric telegraphy in military operations was in the Crimean War in Europe (1854-56), its employment was restricted to communications exchanged among headquarters of the Allies, and some observers were very doubtful about its utility even for this limited usage. It may also be noted that in the annals of that war there is no record of the employment of electric telegraphy together with means for protecting the messages against their interception and solution by the enemy.

On the Union side in the Civil War, military signal operations began with Major Myer's arrival in Washington on 3 June 1861. His basic equipment consisted of kits containing a white flag with a red square in the center for use against a dark background; a red flag with a white square for use against a light background; and torches for night use. It is interesting to note that these are the elements which make up the familiar insignia of our Army Signal Corps. The most pressing need which faced Major Myer was to get officers and men detailed to him wherever signals might be required, and to train them in what had come to be called the "wigwag system,"<sup>1</sup> the motions of which are depicted in Fig. 1. This training included learning something about codes and ciphers, and gaining experience in their usages.

But there was still no such separate entity as a Signal Corps of the

---

<sup>1</sup> And, of course, the G. I.'s of those days had a pet name for the users of the system. They called them "flag floppers."

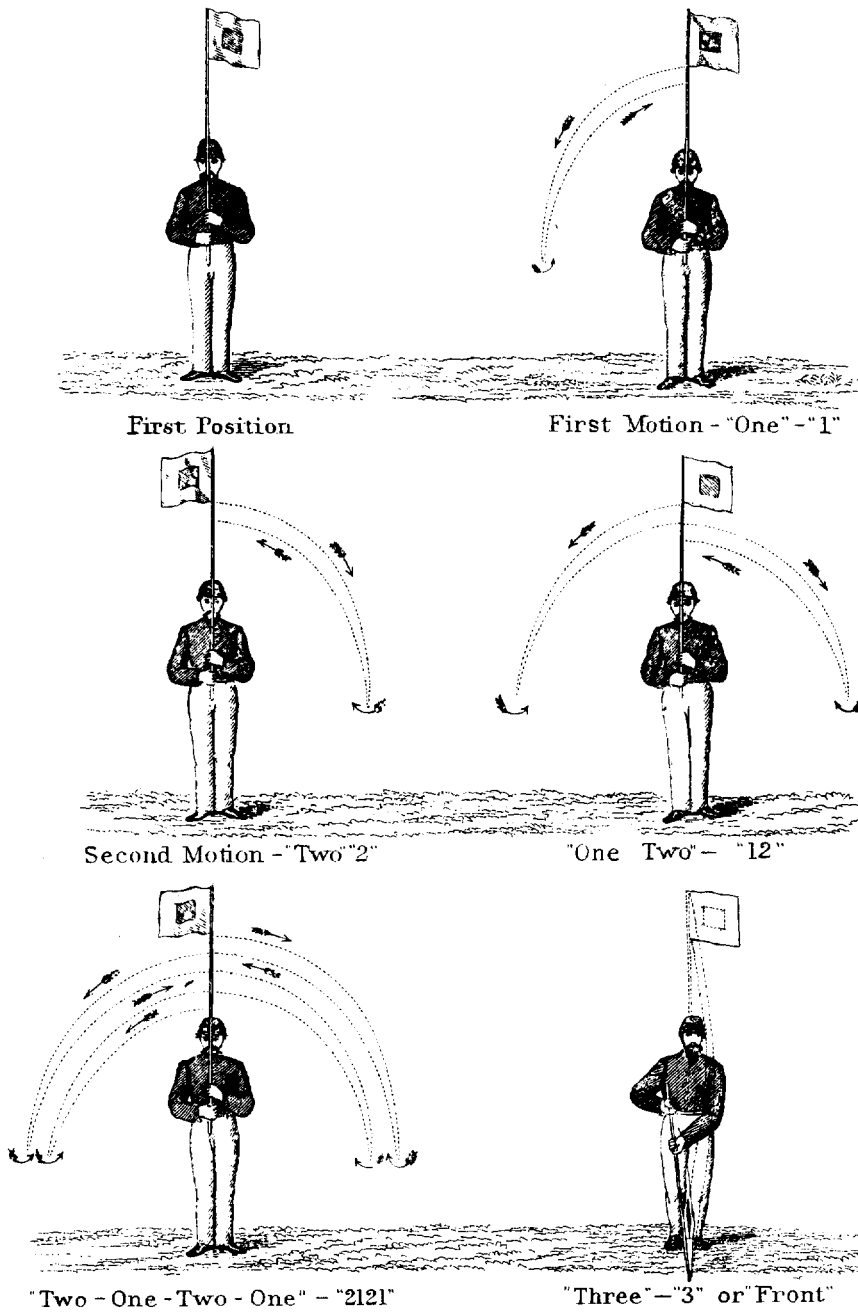
~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

Fig. 1.

~~CONFIDENTIAL~~



W. F. FRIEDMAN

~~CONFIDENTIAL~~

Army. Officers and enlisted men were merely detailed for service with Major Myer for signaling duty. It was not until two years after the war started that the Signal Corps was officially established and organized as a separate branch of the Army, by appropriate Congressional action.

In the meantime, another signaling organization was coming into being—an organization which was an outgrowth of the government's taking over control of the commercial telegraph companies in the United States on 25 February 1862. There were then only three in number: the American, the Western Union, and the Southwestern. The telegraph lines generally followed the right-of-way of the railroads. The then Secretary of War, Simon Cameron, sought the aid of Thomas A. Scott, of the Pennsylvania Railroad, who brought some of his men to Washington for railroad and telegraphic duties with the Federal Government. From a nucleus of four young telegraph operators grew a rather large military telegraph organization which was not given formal status until on 28 October 1861 President Lincoln gave Secretary Cameron authority to set up a "U.S. Military Telegraph Department" under a man named Anson Stager, who, as general superintendent of the Western Union, was called to Washington, commissioned a captain (later a colonel) in the Quartermaster Corps, and made superintendent of the Military Telegraph Department. Only about a dozen of the members of the Department became commissioned officers, and they were made officers so that they could receive and disburse funds and property; all the rest were civilians. The U.S. Military Telegraph "Corps," as it soon came to be designated, without warrant, was technically under Quartermaster General Meigs, but for all practical purposes it was under the immediate and direct control of the Secretary of War, a situation admittedly acceptable to Meigs. There were now two organizations for signaling in the Army, and it was hardly to be expected that no difficulties would ensue from the duality. In fact, the difficulties began very soon, as can be noted in the following extract from a lecture before the Washington Civil War Round Table, early in 1954, by Dr. George R. Thompson, Chief of the Historical Division of the Office of the Chief Signal Officer of the U.S. Army:

The first need for military signals arose at the important Federal fortress in the lower Chesapeake Bay at Fort Monroe. Early in June, Myer arrived there, obtained a detail of officers and men and began schooling them. Soon his pupils were wig-wagging messages from a small boat, directing fire of Union batteries located on an islet in Hampton Roads against Confederate fortifications near Norfolk. Very soon, too, Myer began encountering trouble with commercial wire telegraphers in the area. General Ben Butler, commanding the Federal Department in southeast Virginia, ordered that wire telegraph

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

facilities and their civilian workers be placed under the signal officer. The civilians, proud and jealous of their skills in electrical magic, objected in no uncertain terms and shortly an order arrived from the Secretary of War himself who countermanded Butler's instructions. The Army signal officer was to keep hands off the civilian telegraph even when it served the Army.

I have purposely selected this extract from Dr. Thompson's presentation because in it we can clearly hear the first rumblings of that lengthy and acrimonious feud between two signaling organizations whose uncoordinated operations and rivalry greatly reduced the efficiency of all signaling operations of the Federal Army. As already indicated, one of these organizations was the U.S. Military Telegraph "Corps," hereinafter abbreviated as the USMTC, a civilian organization which operated the existing commercial telegraph systems for the War Department, under the direct supervision of the Secretary of War, Edwin M. Stanton. The other organization was, of course, the infant Signal Corps of the United States Army, which was not yet even established as a separate Branch, whereas the USMTC had been established in October 1861, as noted above. Indeed, the Signal Corps had to wait until March 1863, *two years after the outbreak of war*, before being established officially. In this connection it should be noted that the Confederate Signal Corps had been established a full year earlier, in April 1862. Until then, as I've said before, for signaling duty on both sides, there were only officers who were individually and specifically detailed for such duty from other branches of the respective Armies of the North and the South. Trouble between the USMTC and the Signal Corps of the Union Army began when the Signal Corps became interested in signaling by electric telegraphy and began to acquire facilities therefor.

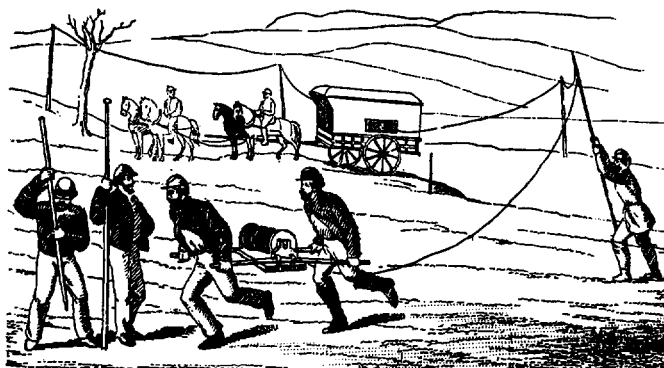
As early as in June 1861, Chief Signal Officer Myer had initiated action toward acquiring or obtaining electrical telegraph facilities for use in the field but with one exception nothing happened. The exception was in the case of the episode in the military department in southeast Virginia, commanded by General Benjamin Butler, an episode that clearly foreshadowed the future road for the Signal Corps in regard to electrical signaling: the road was to be closed and barred. In August 1861, Colonel Myer tried again and in November of the same year he recommended in his annual report that \$30,000 be appropriated to establish an electric signaling branch in the Signal Corps. The proposal failed to meet the approval of the Secretary of War. One telegraph train, however, which had been ordered by Myer many months before, was delivered in January 1862. The train was tried out in an experimental fashion, and under considerable difficulties, the most disheartening of which was the active opposition of persons in Washington, particularly the Secretary of War. So, for

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

practically the whole of the first two years of the war, signal officers on the Northern side had neither electrical telegraph facilities nor Morse operators—they had to rely entirely on the wig-wag system. However, by the middle of 1863 there were thirty “flying-telegraph” trains in use in the Federal Army. Here’s a picture of such a train. The normal length of field telegraph lines was five to eight miles, though in some cases the instruments had worked at distances as great as twenty miles. But even before the Signal Corps began to



A drawing from Myer's *Manual of Signals* illustrating the field, or flying, telegraph. It shows the wagon with batteries and instruments. The wire (in this case presumably bare copper, since it is being strung on insulators on poles) is being run out from a reel carried by two men. The linesmen are using a crowbar to open holes to receive the lance poles. Myer estimated that  $2\frac{1}{2}$  miles of such wire line could be put up in an hour.

Fig. 2.

acquire these facilities, there had been agitation to have them, as well as their Signal Corps operating personnel, all turned over to the USMTC, which had grown into a tightly-knit organization of over 1,000 men and had become very influential in Washington, especially by virtue of its support from Secretary of War Stanton. As a consequence, the USMTC had its way. In the fall of 1863, it took over all the electric telegraph facilities and telegraph operators of the Signal Corps. Colonel Myer sadly wrote: “With the loss of its electric lines the Signal Corps was crippled.”

So now there were two competing signal organizations on the Northern side: The U.S. Army's Signal Corps, which was composed entirely of military personnel with no electric telegraph facilities (but was equipped with means for visual signaling), and the USMTC, which was not a part of the Army, being staffed almost entirely with civilians, and which had electric telegraph facilities and skilled Morse

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

operators (but no means or responsibilities for visual signaling or "aerial telegraphy" which, of course, was old stuff). "Electric telegraphy" was now *the* thing. The USMTC had no desire to share electric telegraphy with the Signal Corps, a determination in which it was most ably assisted by Secretary of War Stanton, for reasons that fall outside the scope of the present lecture.

However, from a technical point of view it is worth going into this rivalry just a bit, if only to note that the personnel of both organizations, the military and the civilian, were not merely signalmen and telegraph operators: they served also as cryptographers and were therefore entrusted with the necessary cipher books and cipher keys. Because of this, they naturally became privy to the important secrets conveyed in cryptographic communications and they therefore enjoyed status as VIP's. This was particularly true of members of the USMTC, because they, and only they, were authorized to be custodians and users of the cipher books. Not even the commanders of the units they served had access to them. For instance, on the one and only occasion when General Grant forced his cipher operator, a civilian named Beckwith, to turn over the current cipher book to a colonel on Grant's staff, Beckwith was immediately discharged by the Secretary of War and Grant was reprimanded. A few days later, Grant apologized and Beckwith was restored to his position. But Grant never again demanded the cipher book held by his telegraph operator.

The Grant-Beckwith affair alone is sufficient to indicate the lengths to which Secretary of War Stanton went to retain control over the USMTC, including its cipher operators, and its cipher books. In fact, so strong a position did he take that on 10 November 1863, following a disagreement over who should operate and control all the military telegraph lines, Myer, by then full Colonel, and bearing the imposing title "Chief Signal Officer of the United States Army," a title he had enjoyed for only two months, was peremptorily relieved from that position and put on the shelf. Not long afterward, and for a similar reason, Myer's successor, Lieutenant Colonel Nicodemus, was likewise summarily relieved as Chief Signal Officer by Secretary Stanton; indeed, he was not only removed from that position—he was "dismissed the Service." Stanton gave "phoney" reasons for dismissing Colonel Nicodemus, but I am glad to say that the latter was restored his commission in March 1865, by direction of the President; also by direction of the President, Colonel Myer was restored to his position as Chief Signal Officer of the U.S. Army on 25 February 1867.

When Colonel Myer was relieved from duty as Chief Signal Officer in November 1863, he was ordered to Cairo, Illinois, to await orders

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

for a new assignment. Very soon thereafter he was either designated (or he may have himself decided) to prepare a field manual on signaling and there soon appeared, with a prefatory note dated January 1864, a pamphlet of 148 pages, a copy of which is now in the Rare Book Room of the Library of Congress. The title page reads as follows:

*"A Manual of Signals: for the use of signal officers in the field. By Col. Albert J. Myer, Signal Officer of the Army, Washington, D. C., 1864."*

Even in this first edition, printed on an Army press, Myer devoted nine pages to a reprint of an article from *Harper's Weekly* entitled "Curiosities of Cipher," and in the second edition, 1866, he expanded the section on cryptography to sixty pages. More editions followed and I think we may well say that Myer's *Manual*, in its several editions, was the pioneer American text on military signaling. But I'm sorry to say that as regards cryptology it was rather a poor thing. Poe had done better twenty years before that in his essay entitled "A few words on secret writing."

Because of its historic nature, you may like to see what Myer's original "wig-wag code" was like. It was called "a two-element code" because it employed only two digits, 1 and 2, in permutations of 1, 2, 3 and 4 groups. For example, A was represented by the permutation 22; B, by 2122; and C, by 121, etc. In flag signaling, a "1" was indicated by a motion to the left, and a "2" by a motion to the right. Later these motions were reversed, for reasons which must have been good but are now not obvious.\* Here is Myer's two-element code which continued to be used until 1912:

## GENERAL SERVICE CODE

A - 22	M - 1221	Y - 111
B - 2122	N - 11	Z - 2222
C - 121	O - 21	& - 1111
D - 222	P - 1212	ing - 2212
E - 12	Q - 1211	tion - 1112
F - 2221	R - 211	
G - 2211	S - 212	End of word - 3
H - 122	T - 2	End of sentence - 33
I - 1	U - 112	End of message - 333
J - 1122	V - 1222	Affirmative - 22.22.22.3
K - 2121	W - 1121	Repeat - 121.121.121
L - 221	X - 2122	Error - 212121

*Note:* No. 3 (end of word) was made by a forward downward motion, called "front". There were about a dozen more signals, for numerals, for frequently used short sentences, etc.

\*This reversal can be seen in Fig. 1.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

We must turn our attention now to the situation as regards the organization for signaling in the Confederate Army. It is of considerable interest to note that in the first great engagement of the War, that of the first Bull Run battle, the Confederate Signal Officer was that young Lieutenant, E. P. Alexander, who had assisted in demonstrating the wig-wag system before a board appointed by the War Department to study Myer's system. Alexander, now a Captain in grey, used Myer's system during the battle, which ended in disaster for the Union forces; and it is said that Alexander's contribution by effective signaling was an important factor in the Confederate victory. Dr. Thompson, whom I have quoted before, says of this battle:

Thus the fortunes of war in this battle saw Myer's system of signals succeed, ironically, on the side hostile to Myer. Because of general unpreparedness and also some disinterest and ignorance, the North had neither wig-wag signals nor balloon observations.

The only communication system which succeeded in signal work for the Union Army was the infant USMTC. But the Confederate system under Alexander, off to a good start at Bull Run, throughout the war operated with both visual and electric telegraphy, and the Confederates thought highly enough of their signal service to establish it on an official basis, on 19 April 1862, less than a year after that battle. Thus, although the Confederate Signal Corps never became a distinct and independent branch of the Army as did the Union Signal Corps, it received much earlier recognition from the Confederate Government than did the Signal Corps of the Federal Government. Again quoting Dr. Thompson:

The Confederate Signal Corps was thus established nearly a year earlier than its Federal counterpart. It was nearly as large, numbering some 1,500, most of the number, however, serving on detail. The Confederate Signal Corps used Myer's system of flags and torches. The men were trained in wire telegraph, too, and impressed wire facilities as needed. But there was nothing in Richmond or in the field comparable to the extensive and tightly controlled civilian military telegraph organization which Secretary Stanton ruled with an iron hand from Washington.

We come now to the codes and ciphers used by both sides in the war, and in doing so we must take into consideration the fact that on the Union side, there were, as I have indicated, two separate organizations for signal communications; one for visual signaling, the other for electric. We should therefore not be too astonished to find that the cryptosystems used by the two competing organizations were different. On the other hand, on the Confederate side, as just noted, there was only one organization for signal communications, the Signal Corps of the Confederate States Army, which used both visual and electric telegraphy, the latter facilities being taken over and employed

~~CONFIDENTIAL~~

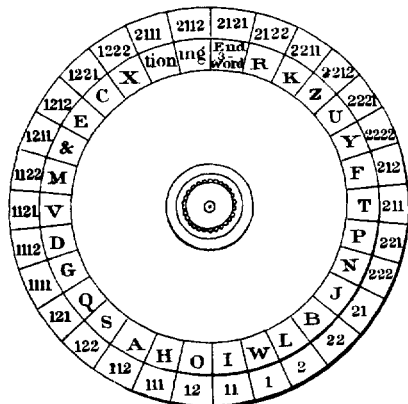
when and where they were available. There were reasons for this marked difference between the way in which the Union and the Confederate signal operations were organized and administered but I do not wish to go into them now. One reason, strange to say, had to do with the difference between the cryptocommunication arrangements in the Union and in the Confederate Armies.

We will discuss the cryptosystems used by the Federal Signal Corps first and then those of the Confederate Signal Corps. Since both corps used visual signals as their primary means, we find them employing Myer's visual-signaling code shown above. At first both sides sent unenciphered messages; but soon after learning that their signals were being intercepted and were being read by the enemy, each side decided to do something to protect its messages. Initially both decided on the same artifice, *viz*, changing the visual-signaling equivalents for the letters of the alphabet, so that, for instance, "22" was not always "A," etc. This sort of changing-about of values soon became impractical, since it prevented memorizing the wig-wag equivalents once and for all. The difficulty in the Union Army's Signal Corps was solved by the introduction into usage of a cipher disk invented by Myer himself. A full description of the disk in its various embodiments will be found in Myer's *Manual*, but here's a picture of three forms of it. You can see how readily the visual wig-wag equivalents for letters, figures, etc., can be changed according to some pre-arranged indicator for juxtaposing concentric disks. In my Fig. 3 the top left disks (Fig. 1 of Myer's Plate XXVI) show that the letter A is represented by 112, B, by 22, etc. By moving the two circles to a different juxtaposition a new set of equivalents will be established. Of course, if the setting is kept fixed for a whole message the encipherment is strictly monoalphabetic; but Myer recommends changing the setting in the middle of the message or, more specifically, at the end of each word, thus producing a sort of polyalphabetic cipher which would delay solution a bit. An alternative way, Myer states, would be to use what he called a "countersign word," but which we call a *keyword*, each letter of which would determine the setting of the disk for a single word or for two consecutive words, etc. Myer apparently did not realize that retaining or showing externally, that is, in the cipher text, the lengths of the words of the plain text very seriously impairs the security of the cipher message. A bit later we shall discuss the security afforded by the Myer disk in actual practice.

In the Confederate Signal Corps, the system used for encipherment of visual signals was apparently the same as that used for enciphering telegraphic messages, and we shall soon see what it was. Although Myer's cipher disk was captured a number of times, it was apparently disdained by the Confederates, who preferred to use a wholly different

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

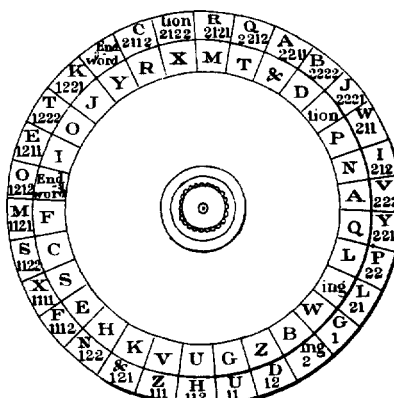
Figure 1.



Two Discs.

Vertical Section

Figure 2.

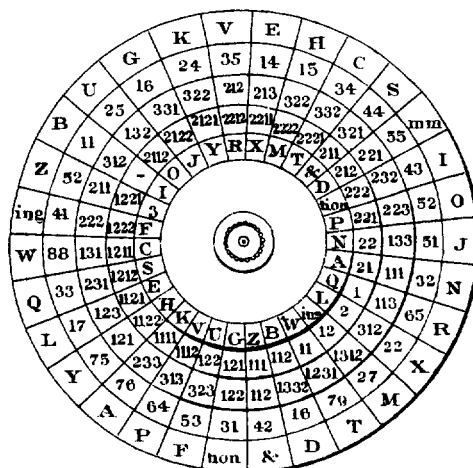


Two Discs.

Figure 3.



Figure 4.



Plan for Service Discs.

Figure 5.



Vertical Section exhibiting plan for four Discs.

Fig. 3.

~~CONFIDENTIAL~~



W. F. FRIEDMAN

~~CONFIDENTIAL~~

type of device, as will be described presently, for both visual and electric telegraphy.

So much for the cryptosystems used in connection with visual signals by the Signal Corps of both the North and the South, systems which we may designate as "tactical ciphers." We come now to the systems used for what we may call "strategic ciphers," because the latter were usually exchanged between the seat of Government and field commanders, or among the latter. In the case of these communications the cryptosystems employed by each side were quite different.

On the Northern side the USMTC used a system based upon what we now call transposition but in contemporary accounts they were called "route ciphers" and that name has stuck. The designation isn't too bad, because the processes of encipherment and decipherment, though dealing not with the individual letters of the message but with entire words, involves following the prescribed paths or routes in a diagram in which the message is written. I know no simpler or more succinct description of the route cipher than that given by one of the USMTC operators, J. E. O'Brien, in an article in *Century Magazine*, XXXVIII, September 1889, entitled "Telegraphing in Battle":

The principle of the cipher consisted in writing a message with an equal number of words in each line, then copying the words up and down the columns by various routes, throwing in an extra word at the end of each column, and substituting other words for important names and verbs.

A more detailed description in modern technical terms would be as follows: A system in which in encipherment the *words* of the plain-text message are inscribed within a matrix of a specified number of rows and columns, inscribing the words within the matrix from left to right, in successive lines and rows downward as in ordinary writing, and taking the words out of the matrix, that is, transcribing them, according to a prearranged route to form the cipher message. The specific routes to be followed were set forth in numbered booklets, each being labeled "War Department Cipher" followed by a number. In referring to them hereinafter I shall use the term "cipher books," or sometimes, more simply, the term "ciphers," although the cryptosystem involves both cipher and code processes. It is true that the basic principle of the system, that of transposition, makes the system technically a cipher system as defined in our modern terminology; but the use of "arbitraries," as they were called, that is, words arbitrarily assigned to represent the names of persons, geographic points, important nouns and verbs, etc., makes the system technically a code system as defined in our modern terminology.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

There were in all about a dozen cipher books used by the USMTC throughout the war. For the most part they were employed consecutively, but, it seems that sometimes two different ones were employed concurrently. They contained not only the specific routes to be used but also indicators for the routes and for the sizes of the matrices; and, of course, there were lists of code words, with their meanings. These route ciphers were supposed to have been the invention of Anson Stager, whom I have mentioned before in connection with the establishment of the USMTC, and who is said to have first devised such ciphers for General McClellan's use in West Virginia, in the summer of 1861, before McClellan came to Washington to assume command of the Army of the Potomac.

Anson Stager and many others thought that he was the original inventor of the system, but such a belief was quite in error because word-transposition methods similar to Stager's were in use hundreds of years before his time. For instance, in 1685, in an unsuccessful attempt to invade Scotland, in a conspiracy to set the Duke of Monmouth on the throne, Archibald Campbell, 9th Earl of Argyll, suffered an unfortunate "accident". He was taken prisoner and beheaded by order of James the Second. The communications of the poor Earl were not secure, and when they fell into government hands they were soon deciphered. The method Argyll used was that of word transposition, and if you are interested in reading a contemporary account of how it was solved, look on pages 56-59 of that little book I mentioned before as being one of the very first books in English dealing with the subject of cryptology, that by James Falconer, entitled *Cryptomenysis Patefacta: Or the Art of Secret Information Disclosed Without a Key*, published in London in 1685. There you will find the progenitor of the route ciphers employed by the USMTC, 180 years after Argyll's abortive rebellion.

The route ciphers employed by the USMTC are fully described in a book entitled *The Military Telegraph during the Civil War*, by Colonel William R. Plum, published in Chicago in 1882. I think Plum's description of them is of considerable interest and I recommend his book to those of you who may wish to learn more about them, but they are pretty much all alike. If I show you one example of an actual message and explain its encipherment and decipherment I will have covered practically the entire gamut of the route ciphers used by the USMTC, so basically very simple and uniform were they. And yet, believe it or not, legend has it that the Southern signalmen were unable to solve any of the messages transmitted by the USMTC. This long-held legend I find hard to believe. In all the descriptions I have encountered in the literature not one of them, save the one quoted above from O'Brien, tries to make these ciphers as simple as

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

they really were; somehow, it seems to me, a subconscious realization on the part of Northern writers, usually ex-USMTC operators, of the system's simplicity prevented a presentation which would clearly show how utterly devoid it was of the degree of sophistication one would be warranted in expecting in the secret communications of a great modern army in the decade 1860-1870, three hundred years after the birth of modern cryptography in the papal states of Italy.

Let us take the plain text of a message which Plum (p. 58) used in an example of the procedure in encipherment. The cipher book involved is No. 4 and I happen to have a copy of it so we can easily check Plum's work. Here's the message to be enciphered:

Washington, D. C.  
July 15, 1863

For Simon Cameron

I would give much to be relieved of the impression that Meade, Couch, Smith and all, since the battle of Gettysburg, have striven only to get the enemy over the river without another fight. Please tell me if you know who was the one corps commander who was for fighting, in the council of war on Sunday night.

(Signed) A. Lincoln

Plum shows the word-for-word encipherment in a matrix of seven columns and eleven rows.<sup>2</sup> He fails to tell us why a matrix of those dimensions was selected; presumably the selection was made at random, which was certainly permissible. (See fig. 4.)

Note the seven "nulls" (non-significant, or "blind" words) at the tops and bottoms of certain columns, these being added to the cipher text in order to confuse a would-be decipherer. At least that was the theory, but how effective this subterfuge was can be surmised, once it became known that employing nulls was the usual practice. Note also the two nulls (*bless* and *him*) at the end of the last line to complete that line of the matrix. Words in italics are "arbitraries" or code words.

The cipher message is then copied down following the route prescribed by the indicator "BLONDE," as given on page 7 of Cipher Book No. 4 for a message of 11 lines. The indicator could have also been "LINIMENT."

<sup>2</sup> Ruled paper was provided to aid in accuracy. In the diagram the upper of each pair of lines of writing is the cipher, the lower one, the plain text. Simon Cameron was Lincoln's Secretary of War until Jan 1862, when he was replaced by Edwin M. Stanton. If this message cited by Plum is authentic, and there is no reason to doubt this, then Cameron was still in friendly contact with Lincoln, possibly as a special observer.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

1	2	3	4	5	6	7
(heavy)				(county)	(square)	
(null)				(null)	(null)	
<i>Incubus</i>	<i>Stewart</i>	<i>Brown</i>	<i>Norris</i>	<i>Knox</i>	<i>Madison</i>	
Wash., D.C.	July	15th	18	60	3	for
sigh	man	Cammer	on	flea	I	wood
Simon		Cameron		(period)	I	would
give	much	Toby	<i>trammeled</i>	<i>serenade</i>	impression that	
give	much	to be	relieved	of the	impression that	
<i>Bunyan</i>	<i>bear</i>	<i>ax</i>	<i>cat</i>	<i>children</i>	and	awl
Meade	, (comma)	Couch	, (comma)	Smith	and	all
<i>bat</i>	since	the	<i>knit</i>	of	get	ties
, (comma)	since	the	battle	of	Gettys	
<i>large</i>	ass	have	striven	only	to	get
burg	, (comma)	have	striven	only	to	get
<i>village</i>	<i>skeleton</i>	<i>turnip</i>	without	another	<i>optic</i>	<i>hound</i>
the enemy	over	the river	without	another	fight	(period)
Please	tell	me	if	you	no	who
Please	tell	me	if	you	know	who
was	the	<i>Harry</i>	<i>Madrid</i>	<i>locust</i>	who	was
was	the	one	corps	commander	who	was
for	<i>oppressing</i>	<i>bitch</i>	<i>quail</i>	<i>counsel</i>	of	war
for	fighting	, (comma)	in the	council	of	war
on	<i>Tyler</i>	<i>Rustle</i>	<i>upright</i>	<i>Adrian</i>	bless	him
on	Sunday	night	Signature	A. Lincoln	(null)	(null)
	(monkey)	(silk)	(martyr)			(suicide)
	(null)	(null)	(null)			(null)

Fig. 4.

		8		7		4		2	
8			16	11		11			
		13		11		9			
6			5	1					

Bedroom.....	1	.....	Levy
Bedstead.....	2	.....	League
Beverage.....	3	.....	Leather
Beyond.....	4	.....	Legacy
Big.....	5	.....	Lemon
Bill.....	6	.....	Lesson
Billiards.....	7	.....	Lot
Billion.....	8	.....	Library
Blanket.....	9	.....	Life
Bliss.....	10	.....	Lisan
Blonde.....	11	.....	Litiment

Fig. 5.

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

To explain the diagram at the top of Fig. 5 I will show you the "Directions for Use" which appear on the reverse side of the title page of "War Department Cipher No. 4," because I'm afraid you wouldn't believe me if I merely told you what they say. In Fig. 6 is a picture of the title page and I follow it with Fig. 7, a photograph of what's on its reverse.

## DIRECTIONS FOR USE.

## WAR DEPARTMENT CIPHER NO. 4.

To find the route, read the figures in the table at top of page from left to right in the order that they occur alternately in the upper and lower lines, the two intermediate lines of figures having no connection with the route, being introduced simply as a blind, the upper line of figures denoting the route *down* the column and the lower line *up*.

## EXAMPLE.

See page 14; 7 columns.

Route—Up the 3d; down the 6th; up the 1st; down the 7th; up the 2d; down the 4th; up the 5th.

Commence a cipher with one of the "line indicators," taken from same page as route used, which word must indicate the number of lines in the message. Use two words for more than twenty lines.

*Used by the Federal Army in the Civil War*  
3.

Fig. 6.

Fig. 7.

Do you imagine that the chap who was responsible for getting this cipher book approved ever thought about what he was doing when he caused those "Directions for Use" to be printed? It doesn't seem possible. All he would have had to ask himself was, "Why put this piece of information in the book itself?" Cipher books before this have been captured. Suppose this one falls into enemy hands; can't he read, too, and at once learn about the intended deception? Why go to all the trouble of including "phoney" routes anyway? If the book doesn't fall into enemy hands what good are the "phoney" routes anyway? Why not just indicate the routes in a straightforward manner, as had been done before? Thus: "Up the 6th column (since "6" is the first number at the left of the diagram), down the 3rd, up the 5th, down the 7th, up the 1st, down the 4th and down the 2nd." This matter is so incredibly fatuous that it is hard to understand how sensible men—and they were sensible—could be so illogical in their thinking processes. But there the "Directions for Use" stand, for all the world to see and to judge.

Now for the transposition step. The indicator "BLONDE" signifies a matrix of seven columns and eleven rows, with the route set

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

forth above, viz, up the 6th column, down the 3rd, etc., so that the cipher text with a "phoney" address and signature,<sup>3</sup> becomes as follows:

TO A. HARPER CALDWELL,

Washington, D. C.

Cipher Operator, Army of the Potomac:

Blonde bless of who no optic to get and impression I Madison square  
Brown cammer Toby ax the have turnip me Harry bitch rustle silk  
Adrian counsel locust you another only of children serenade flea Knox  
County for wood that awl ties get hound who was war him suicide on  
for was please village large bat Bunyan give sigh incubus heavy Norris  
on trammled cat knit striven without if Madrid quail upright martyr  
Stewart man much bear since ass skeleton tell the oppressing Tyler  
monkey.

(Signed) D. HOMER BATES

Note that the text begins with the indicator "BLONDE". In decipherment the steps are simply reversed. The indicator tells what size matrix to outline; the words beginning "bless of who no optic . . ." are inscribed within the matrix: up the 6th column; then, omitting the "check word" or "null" (which in this case is the word "square") down the 3rd column, etc. The final result should correspond to what is shown in Fig. 4. There then follows the step of interpreting orthographic deviations, such as interpreting "sigh", "man," "cammer," and "on" as Simon Cameron; the word "wood" for "would", etc. The final step reproduces the original plain text.

Save for one exception, all the route ciphers used by the USMTC conformed to this basic pattern. The things that changed from one cipher book to the next were the indicators for the dimensions of the matrices and for the routes, and the "arbitraries" or code equivalents for the various items comprising the "vocabulary," the number of them increasing from one edition to the next, just as might be expected. The sole exception to this basic pattern is to be seen in Cipher Book No. 9 and on only one page of the book. I will show you that page. (See fig. 8.)

What we have here is a deviation from the straightforward route transposition, "up the . . . column, down the . . . column," etc. By introducing one diagonal path in the route (the 6th, 7th, 8th, 9th, 10th words in a message of five columns, and the 1st, 2nd, 3rd, 4th, 5th, and 6th words in a message of six columns) the simple up and down route no longer holds true. The words on the diagonal interrupt the normal up and down paths and introduce complexities in

<sup>3</sup> It was the usual practice to use for address and signature the names of the USMTC operators concerned.

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

12

Message or Division of ...6... Lines.

COMMENCEMENT WORDS.

Yates } .....5..... Stanton } .....6..... Halleck } .....7.....  
 Lincoln } ..... COLUMNS } ..... Buell } .....  
 Chase } ..... McDowell } ..... Sibley } .....  
*Seven* Route.—Up the ...4... column; down the...3...; up the...5...; down the...2...;  
 up the...1...; down the...6...; up the...7...

*Five columns.*

15	25	26	16	6
14	24	27	7	5
13	23	8	17	4
12	9	28	18	3
10	22	29	19	2
11	21	30	20	1

*Six columns.*

6	17	27	36	26	16
7	5	28	35	25	15
8	18	4	34	24	14
9	19	29	3	23	13
10	20	30	33	2	12
11	21	31	32	22	1

Fig. 8.

the method. In fact, the complexities, seemed to be a bit too much for the USMTC cipher operators because, as far as available records show, these complicated routes were never used.

I now wish to make a number of general and a few specific comments on Plum's description of the cryptosystems used by the USMTC.

First, we have learned that although Anson Stager has been credited with inventing the type of cipher under consideration in this study, he was anticipated in the invention by about 200 years. Also, he is given the lion's share of the credit for devising those ciphers although he did have a number of collaborators. Plum names four of them, presumably because he thought them worthy of being singled out for

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

particular attention. Plum and others tell us that copies of messages handled by the USMTC were sometimes intercepted by the enemy but not solved. He cites no authority for this last statement, merely saying that such intercepts were published in the newspapers of the Confederacy with the hope that somebody would come up with their solution. And it may be noted that none of the Confederate accounts of war activities cite instances of the solution of intercepted USMTC messages, although there are plenty of citations of instances of interception and solution of enciphered visual transmissions of the Federal Army's Signal Corps.

Plum states that 12 different cipher books were employed by the Telegraph Corps, but I think there were actually only eleven. The first one was not numbered, and this is good evidence that a long war was not expected. This first cipher book had 16 printed pages. But for some reason, now impossible to fathom, the sequence of numbered books thereafter was as follows: Nos. 6 and 7, which were much like the first (unnumbered) one; then came Nos. 12, 9, 10—in that strange order; then came Nos. 1 and 2; finally came Nos. 3, 4, and 5. (Apparently there was no No. 8, or No. 11—at least they are never mentioned.) It would be ridiculous to think that the irregularity in numbering the successive books was for the purpose of communication security, but there are other things about the books and the cryptosystem that appear equally silly. There may have been good reasons for the erratic numbering of the books, but if so, what they were is now unknown. Plum states that No. 4, the last one used in the war, was placed into effect on 23 March 1865, and that it and all other ciphers were discarded on 20 June 1865. However, as noted, there was a No. 5, which Plum says was given a limited distribution. I have a copy of it, but whether it was actually put into use I do not know. Like No. 4, it had 40 pages. About 20 copies were sent to certain members of the USMTC, scattered among 12 states; and, of course, Washington must have had at least one copy.

We may assume with a fair amount of certainty that the first (the unnumbered) cipher book used by the USMTC was merely an elaboration of the one Stager produced for the communications of the governors of Ohio, Indiana and Illinois, and of which a copy is given by only one of the writers who have told us about these ciphers, namely, David H. Bates. Bates, in his series of articles entitled "Lincoln in the Telegraph Office" (*The Century Magazine*, Vol. LXXIV, Nos. 1-5, May-Sept, 1907)<sup>4</sup> shows a facsimile thereof (p.

<sup>4</sup> The series was then put out in book form under the same title by the D. Appleton-Century Company, New York, 1907, reprinted in 1939.

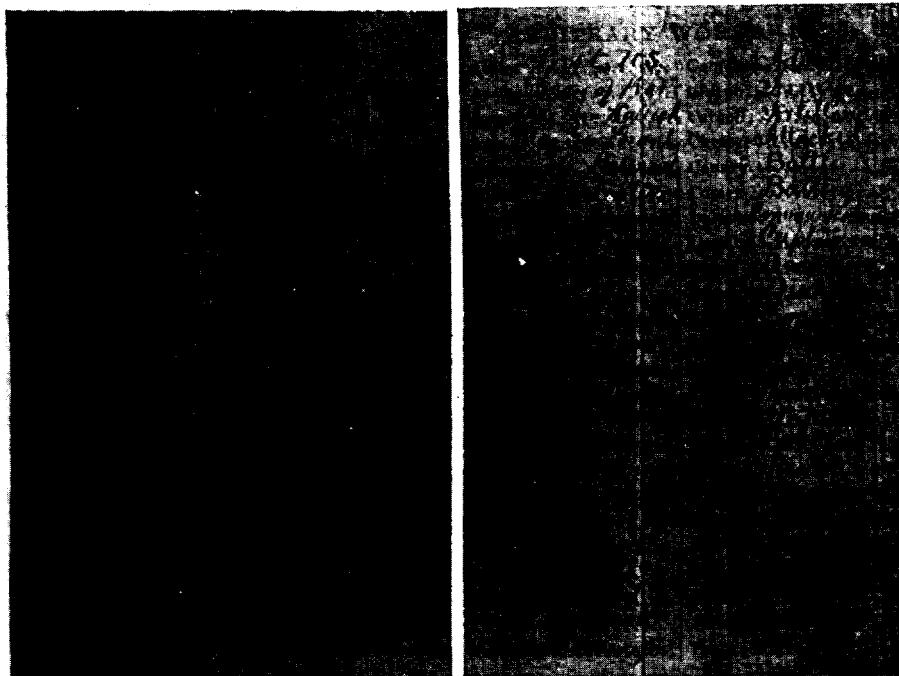
~~CONFIDENTIAL~~



W. F. FRIEDMAN

~~CONFIDENTIAL~~

292, June 1907 issue), and I have had as good a reproduction made of it as is possible from the rather poor photographic facsimile. The foregoing cipher is the prototype upon which all subsequent cipher books were based, the first of the War Department series being the one shown by Plum.



FACSIMILE OF THE TELEGRAPHIC CIPHER-CODE USED BY THE UNITED STATES GOVERNMENT IN 1861

Fig. 9.

When these ciphers came into use it was not the practice to misspell certain words intentionally; but as the members of the USMTC (who, as I've told you, not only served as telegraph operators but also as cipher clerks) developed expertness, the practice of using non-standard orthography was frequently employed to make solution of messages more difficult. You have already seen examples of this practice, and one can find hundreds of other examples of this sort of artifice. Then, further to increase security, more and more code equivalents were added to represent such things as ordinal and cardinal numbers, months of the year, days of the week, hours of the day, punctuation, etc. As a last step, additional code equivalents for frequently used words and phrases were introduced. One good example of two typical pages from one of these books will characterize them all.

You will notice that the code equivalents are printed but their meanings are written in by hand. This was usually the case, and the reason is obvious: for economy in printing costs, because the printed code equivalents of plaintext items in cipher books belonging to the same series are identical; only their meanings change from one book to another, and of course, the transposition routes, their indicators, and other variables change from one book to another. I am fortunate in having six of these cipher books in my private collection, so that comparisons among them are readily made. The first feature to be noted is that the code equivalents are all good English dictionary words (or proper nouns), of not less than three nor more than seven (rarely eight) letters. A careful scrutiny shows that in the early editions the code equivalents are such as are not very likely to appear as words in the plaintext messages; but in the later editions, beginning with No. 12, *more than 50% of the words used as code equivalents are such as might well appear in the plaintext of messages.* For example, words such as AID, ALL, ARMY, ARTILLERY, JUNCTION, CONFEDERATE, etc., baptismal names of persons, and names of cities, rivers, bays, etc., appear as code equivalents. Among names used as code equivalents are SHERMAN, LINCOLN, THOMAS, STANTON, and those of many other prominent officers

W. F. FRIEDMAN

~~CONFIDENTIAL~~

and officials of the Union Army and the Federal Government, as well as of the Confederate Army and Government; and, even more intriguing, such names were employed as indicators for the number of columns and the routes used—the so-called “Commencement Words.” It would seem that names and words such as those I’ve mentioned might occasionally have brought about instances where difficulty in deciphering messages arose from this source of confusion, but the literature doesn’t mention them. I think you already realize why such commonly-used proper names and words were not excluded. There was, indeed, method in this madness.

But what is indeed astonishing to note is that in the later editions of these cipher books, in a great majority of cases the words used as “arbitraries,” differ from one another by at least two letters (for example, LADY, and LAMB, LARK and LAWN, ALBA and ASIA, LOCK and WICK, MILK and MINT), or by more than two (for example MYRTLE and MYSTIC, CARBON and CANCER, ANDES and ATLAS). One has to search for cases in which two words differ by only one letter, but they can be found if you search long enough for them, as, for example, QUINCY and QUINCE, PINE and PIKE, NOSE and ROSE. Often there are words with the same initial trigraph or tetragraph, but then the rest of the letters are such that errors in transmission or reception would easily manifest themselves, as, for example, in the cases of MONSTER and MONARCH, MAGNET and MAGNOLIA. All in all, it is important to note that the compiler or compilers of these cipher books had adopted a principle known today as the “two-letter differential,” a feature found only in codebooks of a much later date. In brief, the principle involves the use, in a given codebook, of code groups differing from one another by at least two letters. This principle is employed by knowledgeable code compilers to this very day, not only because it enables the recipient of a message to detect errors in transmission or reception, but also to correct them. This is made possible if the permutation tables used in constructing the code words are printed in the codebooks, so that most errors can be corrected without calling for a repetition of the transmission. It is clear, therefore, that the compilers of these cipher books took into consideration the fact that errors are to be expected in Morse telegraphy, and by incorporating, but only to a limited extent, the principle of the two-letter differential, they tried to guard against the possibility that errors might go undetected. Had artificial 5-letter groups been used as code equivalents, instead of dictionary words, possibly the cipher books would also have contained the permutation tables. But it must be noted that permutation tables made their first appearance only about a quarter of a century after the Civil

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

War had ended, and then only in the most advanced types of commercial codes.

There is, however, another feature about the words the compilers of these books chose as code equivalents. It is a feature that manifests real perspicacity on their part, and you probably already have divined it. A few moments ago I said that I would explain why, in the later and improved editions of these books, words which might well be words in plaintext messages were not excluded from the lists of code equivalents: it involves the fact that the basic nature of the cryptosystem in which these code equivalents were to be used was clearly recognized by those who compiled the books. Since the cryptosystem was based upon *word* transposition, what could be more confusing to a would-be cryptanalyst, working with messages in such a system, than to find himself unable to decide whether a word in the cipher text of a message he is trying to solve is actually in the original plaintext message and has its normal meaning, or is a code word with a secret significance—or even a null, a non-significant word, a “blind” or a “check word,” as those elements were called in those days? That, no doubt, is why there are, in these books, so many code equivalents which might well be “good” words in the plaintext messages. And in this connection I have already noted an additional interesting feature: at the top of each page devoted to indicators for signaling the number of columns or rows in the specific matrix for a message are printed the so-called “commencement words,” or what we now call “indicators”. Now there are nine such words, in sets of three, any one of which *could* actually be a real word or name in the plaintext message. Such words when used as indicators could be very confusing to enemy cryptanalysts, especially after the transposition operation. Here, for example, are the “commencement words” on page 5 of cipher book No. 9: Army, Anson, Action, Astor, Advance, Artillery, Anderson, Ambush, Agree; on page 7 of No. 10: Cairo, Curtin, Cavalry, Congress, Childs, Calhoun, Church, Cobb, etc. Moreover, in Nos. 1, 3, 4, 5, and 10 the “line indicators,” that is, the words indicating the number of horizontal rows in the matrix, are also words such as could easily be words in the plaintext messages. For example, in No. 1, page 3, the line indicators are as follows:

Address	1	Faith	Assume	6	Bend
Adjust	2	Favor	Awake	7	Avail
Answer	3	Confine	Encamp	8	Active
Appear	4	Bed	Enroll	9	Absent
Appeal	5	Beef	Enough	10	Accept

Note two things in the foregoing list: first, there are variants—

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

there are two indicators for each case; and second, the indicators are not in strict alphabetic sequence. This departure from strict alphabeticity is even more obvious in the pages devoted to vocabulary, a fact of much importance cryptanalytically. Note this feature, for example, in Fig. 10, which shows pages 14 and 15 of cipher book No. 12.

In this respect, therefore, these books partake somewhat of the nature of two-part or "randomized" codes, or, in British terminology, "hatted" codes. In the second lecture of this series the physical difference between one-part and two-part codes was briefly explained, but an indication of the technical cryptanalytic difference between these two types of codes may be useful at this point. Two-part codes are much more difficult to solve than one-part codes, in which both the plaintext elements and their code equivalents progress in parallel sequences. In the latter type, determination of the meaning of one code group quickly and rather easily leads to the determination of the meanings of other code groups above or below the one that has been solved. For example, in the following short but illustrative example, if the meaning of code group 1729 has been determined to be "then", the meaning of the code group 1728 could well be "the" and that of

1728 — the	7621 — the
1729 — then	0972 — then
1730 — there	1548 — there

the code group 1730, "there". But in a two-part code, determining the meaning of the code group 0972 to be "then" gives no clue whatever as to the meaning of the groups 7621 or 1548. For ease in decoding messages in such a code there must be a section in which the code groups are listed in numerical sequence, and are accompanied by their meanings, which, of course, will be in a random sequence. The compilers of the USMTC cipher books must have had a very clear idea of what I have just explained, but they made a compromise of a practical nature between a strictly one-part and a strictly two-part code, because they realized that a code of the latter sort is twice as bulky as one of the former sort, besides being much more laborious to compile and check the contents for accuracy. The arrangement they chose wasn't too bad, so far as cryptosecurity was concerned. As a matter of fact, and speaking from personal experience in decoding a rather long message addressed to General Grant, I had a difficult time in locating many of the code words in the book, because of the departure from strict alphabeticity. I came across that message in a workbook in my collection, the workbook of one of the important members of the USMTC—none other than our friend

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

Plum, from whose book, *The Military Telegraph during the Civil War*, comes much of the data I've presented in this lecture. On the fly-leaf of Plum's workbook there appears, presumably in his own handwriting, the legend "W. R. Plum Chf Opr with Gen. G. H. Thomas". Here's one of the messages he enciphered in cipher book No. 1, the book in which, he says, more important telegrams were sent than in any other:

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

Fig. 11.

Note how many "arbitrariness" appear in the plaintext message, that is *before* transposition. After transposition the melange of plaintext, code words, indicators and nulls makes the cryptogram mystifying.<sup>5</sup> And yet, was the system as inscrutable as its users apparently thought? It is to be remembered, of course, that messages were then transmitted by wire telegraphy, not by radio, so that enemy messages could be obtained only by "tapping" telegraph lines or capturing couriers or headquarters with their files intact. Opportunities for these methods of acquiring enemy traffic were not frequent, but they did occur from time to time, and in one case a Confederate signalman hid in a swamp for several weeks and tapped a Federal telegraph line, obtaining a good many messages. What success, if any, did Confederate cryptanalysts have in their attempts to solve such USMTC cryptograms as they did intercept? We shall try to answer this question in due time.

<sup>5</sup> In searching for a good example my eye caught the words "Lincoln shot" at the left of the matrix and I immediately thought that the message had to do with Booth's assassination of the President. But after hurriedly translating the message and finding nothing in it having anything to do with the shooting it occurred to me to look up the indicators for a matrix of six rows and eight columns. They turned out to be LINCOLN (message of 8 columns), SHOT (6 rows). The word SMALL beneath the "Lincoln shot" is a variant for SHOT, also meaning "6 rows".

W. F. FRIEDMAN

~~CONFIDENTIAL~~

As indicated earlier, there were no competing signal organizations in the Confederacy as there were on the Union side. There was nothing at the center of government in Richmond or in the combat zone comparable to the extensive and tightly-controlled civilian military telegraph organization which Secretary Stanton ruled with such an iron hand from Washington. Almost as a concomitant, it would seem, there was in the Confederacy, save for two exceptional cases, one and only one officially-established cryptosystem to serve the need for protecting tactical as well as strategic communications, and that was the so-called Vigenère Cipher, which apparently was the cipher authorized in an official manual prepared by Captain J. H. Alexander as the partial equivalent of Myer's *Manual of Signals*. You won't find the name Vigenère in any of the writings of contemporary signal officers of either the North or the South. The signalmen of those days called it the "Court Cipher", this term referring to the system in common use for diplomatic or "court" secret communications about this period in history. It is that cipher which employs the so-called Vigenère Square with a repeating key.<sup>6</sup> Here is the square which Plum calls the "Confederate States Cipher Key" and which is followed by his description of its manner of employment. (See figs. 12a and 12b.)

There are certain comments to be made on the sample messages. In the first place, note that in the first message certain words are left unenciphered; in the second place, in both the first and second messages, the ciphers retain and clearly show the lengths of the words which have been enciphered. Both of these faulty practices greatly weaken the security of ciphers because they leave good clues to their contents and can easily result in facilitating solution of the messages. We know today that cipher messages must leave nothing in the clear. Even the address and the signature, the date, time and place of origin, etc., should if possible be hidden; and the cipher text should be in completely regular groupings, first, so as not to disclose the lengths of the plaintext words, and second, to promote accuracy in transmission and reception.

So far as my studies have gone, I have not found a single example of a Confederate Vigenère cipher which shows neither of these two fatal weaknesses. The second of the two examples is the only case I have found in which there are no unenciphered words in the text of the message. And the only example I have been able to find in

---

<sup>6</sup> A keyword is employed to change the alphabets cyclically, thus making the cipher what is called today a periodic or multiple-alphabet cipher controlled by the individual letters of a key, which may consist of a word, a phrase, or even of a sentence, repeated as many times as necessary.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

38

## THE MILITARY TELEGRAPH DURING THE

## CONFEDERATE STATES CIPHER KEY.

	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
1	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
2	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
3	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
4	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
5	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
6	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
7	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
8	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
9	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
10	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
11	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
12	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
13	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
14	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
15	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
16	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
17	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
18	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
19	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
20	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
21	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
22	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
23	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
24	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
25	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
26	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 Words.*—Complete Victory. Manchester Bluff.

To put into cipher the first message, which is put up by using "Manchester Bluff" as the key, and the second by the key term, "Complete Victory," find at the left-hand side of the table the first letter of the first word to be ciphered, and at the top of the table, the first letter of the key term. At the junction of the columns in which these letters are so found, will be seen the arbitrary letter which is to be used in lieu of the real one at the left. Continue in this way with each successive letter of the message and key term, repeating on the latter till finished. Thus, "Sherman is victorious," put in cipher by using the first key, would read, as shown by the capitals, <sup>C-o-m-p-l-e-t-e-v-i-c-t-o-</sup>UVQG XEG MN DKVH

Fig. 12-A.

~~CONFIDENTIAL~~



W. F. FRIEDMAN

~~CONFIDENTIAL~~

r-y. C-o-m-p-  
F P K C G H. Of course, any change in the key word, term or phrase changes the arbitrariness, and if neither the real message nor the key is known, it would be somewhat vexatious working it out, unless there were some such suggestive words as occur in Davis's message above, which indicate the ciphered words very clearly; e. g., "By which you may effect" <sup>o t p q g e x y k</sup> a crossing "above that part" <sup>h j o p g k w m e t</sup> of the river. This meaning occurred to the author, at first sight, and doubtless would to any one familiar with military affairs in that section. Having guessed real words, it is very easy to work out the letters of the key. The following two important ciphers were transmitted as divided below; i. e., each word was sent separately, not all mixed, as in the Pemberton cipher. This division does not facilitate translation by the key at all, but materially assists without it, and was, therefore, bad practice. We give below, each message, with its translation, because these telegrams were very important. The curious reader may, at his leisure, by using the key board, study out the key terms, one of which will be found entirely new and quite apropos, in the light of what speedily followed.

CONFEDERATE STATES OF AMERICA, MILITARY TELEGRAPH. Dated  
Head-quarters, February 25, 1865. Received at Richmond, Va.,  
12:25 minutes, A. M.

To HON. J. C. BRECKENRIDGE, Sec'y of War:—I recommend  
that the tysmee fm goutwp rlatvmp ubwaqbtm exfvxj and iswaqjru  
ktmtl are not of immediate necessity, uv kpgfmbpgr mpc thnlfl  
should be lughtsp. (Signed) R. E. LEE.

TRANSLATION.—I recommend that the removal of public property,  
machinery, stores and archives which are not of immediate necessity,  
be commenced. All powder should be secured.

HEAD-QUARTERS C. S. ARMIES, March 24, 1865.

GEN. E. KIRBY SMITH, comdg. Trans-Miss. Dept., Gen.:—Vvq  
ecilmympm rvcog ui lhonnides kfeli kdf wasptf us tfcfsto abxc  
bjx azjkhmgjsiimivbceq qb ndel ueisu ht kfg auld egh opcm mfs  
uvajwh xrymcoci yu dddxtmpt iu icjqkpxt es vvjau mvrw twhte abxc  
iu eoieg o rdegx en uer pv ntiptyxec rqvaryyb rgzq rspz rksjcpb ptax  
rsp ekez raecdstrzpt mzmseb acgg nsfqvfv mc kfg snhe ftrf wh  
mvv kkgge pyh fefm ckfrlisityxl xj jtbbx rq htxd wbbz awvv fd acgg  
avxwzv yciag oe nzy fet lgxa scuh.

I am most respectfully your obdt. servt.,

(Signed) R. E. LEE.

TRANSLATION.—Gen.: The President deems it advisable that  
you should be charged with the military operations on both banks  
of the Miss., and that you should endeavor as promptly as possible to  
cross that river with as large a force as may be prudently withdrawn  
from your present Dept. You will accordingly extend your command  
to the east bank of the Miss., and make arrangements to bring to  
this side such of your present force as you may deem best.

I am most respectfully your obedient servant.

Fig. 12-B

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

which word lengths are not shown (save for one word) is in the case of the following message:

Vicksburg, Dec. 26, 1862.

GEN. J. E. JOHNSTON, JACKSON:

I prefer oaavvr, it has reference to xhvkjqchffabpzelreqpzwnyk to prevent anuzeyxswstpjw at that point, raeelpsgghvelvtzfautlilast lhifnaigtsmmmlfgccajd.

(Signed) J. C. PEMBERTON  
Lt. Gen. Comdg.

Even in this case there are unenciphered words which afforded a clue which enabled our man Plum to find the key and solve the message. It took some time, however, and the story is worth telling.

According to Plum, the foregoing cipher message was the very first one captured by USMTC operators, and it was obtained during the siege of Vicksburg, which surrendered on 4 July 1863. But note the date of the message: 26 December 1862. What was done with the captured message during the months from the end of December 1862 to July 1863? Apparently nothing. Here is what Plum reports:

What efforts General Grant caused to be made to unravel this message, we know not. It was not until October, 1864, that it and others came into the hands of the telegraph cipherers, at New Orleans, for translation . . . .

The New Orleans operators who worked out this key (Manchester Bluff) were aided by the Pemberton cipher and the original telegram, which was found among that general's papers, after the surrender of Vicksburg; also by the following cipher dispatch, and one other.

Plum gives the messages involved, their solution, and the keys, the latter being the three cited above. It would seem that if the captured Pemberton message had been brought to General Grant's attention and he did nothing about it, he was not much interested in intelligence. Secondly, the solution of the Pemberton message and the others apparently took some time, even though there was one message with its plain text (the Pemberton message) and two messages not only with interspersed plaintext words but also with spaces showing word lengths. But Plum does not indicate how long it took for solution. Note that he merely says that the messages came into the hands of the telegraph cipherers in October 1864; he does not tell when solution was reached.

In the various accounts of these Confederate ciphers there is one and only one writer who makes a detailed comment on the two fatal practices to which I refer. A certain Dr. Charles E. Taylor, a Confederate veteran (in an article entitled "The Signal and Secret Service of the Confederate States", published in the Confederate Veteran,

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

Vol. XL, Aug-Sept 1932), after giving an example of encipherment according to the "court cipher" says:

It hardly needs to be said that the division between the words of the original message as given above was not retained in the cipher. Either the letters were run together continuously or breaks, as if for words, were made at random. Until the folly of the method was revealed by experience, only a few special words in a message were put into cipher, while the rest was sent in plain language. Thus . . . I think it may be said that it was impossible for well prepared cipher to be correctly read by any one who did not know the key-word. Sometimes, in fact, we could not decipher our own messages when they came over telegraph wires. As the operators had no meaning to guide them, letters easily became changed and portions, at least, of messages rendered unmeaningly (sic) thereby.

Frankly, I don't believe Dr. Taylor's comments are to be taken as characterizing the practices that were usually followed. No other ex-signalman who has written about the ciphers used by the Confederate Signal Corps makes such observations and I think we must simply discount what Dr. Taylor says in this regard.

It would certainly be an unwarranted exaggeration to say that the two weaknesses in the Confederate cryptosystem cost the Confederacy the victory for which it fought so mightily, but I do feel warranted at this moment in saying that further research may well show that certain battles and campaigns were lost because of insecure cryptocommunications.

A few moments ago I said that, save for an exception or two, there was in the Confederacy one and only one cryptosystem to serve the need for secure tactical as well as strategic communications. One of these exceptions concerned the cipher used by General Beauregard after the battle of Shiloh (8 April 1862). This cipher was purely monoalphabetic in nature and was discarded as soon as the official cipher system was prescribed in Alexander's manual. It is interesting to note that this was done after the deciphered message came to the attention of Confederate authorities in Richmond via a northern newspaper. It is also interesting to note that the Federal War Department had begun using the route cipher as the official system for USMTC messages very promptly after the outbreak of war, whereas not until 1862 did the Confederate States War Department prepare an official cryptosystem, and then it adopted the "court cipher."

The other exception involved a system used at least once before the official system was adopted and it was so different from the latter that it should be mentioned. On 26 March 1862, the Confederate States President, Jefferson Davis, sent General Johnston by

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

special messenger a dictionary, with the following accompanying instruction:<sup>7</sup>

I send you a dictionary of which I have the duplicate, so that you may communicate with me by cipher, telegraphic or written, as follows: First give the page by its number; second, the column by the letter L, M or R, as it may be, in the left-hand, middle, or right-hand columns; third, the number of the word in the column, counting from the top. Thus, the word junction would be designated by 146, L, 20.

The foregoing, as you no doubt have already realized, is one of the types of cryptosystems used by both sides during the American Revolutionary Period almost a century before, except that in this case the dictionary had three columns to the page instead of two. I haven't tried to find the dictionary but it shouldn't take long to locate it, since the code equivalent of the word "junction" was given: 146, L, 20. Moreover, there is extant at least one fairly long message, with its decode. How many other messages in this system there may be in National Archives I don't know.

Coming back now to the "court cipher," you will probably find it just as hard to believe, as I find it, that according to all accounts three and only three keys were used by the Confederates during the three and a half years of warfare from 1862 to mid-1865. It is true that Southern signalmen make mention of frequent changes in key but only the following three are specifically cited:

- 1) COMPLETE VICTORY
- 2) MANCHESTER BLUFF
- 3) COME RETRIBUTION.

It seems that all were used concurrently. There may have been a fourth key, IN GOD WE TRUST, but I have seen it only once, and that is in a book explaining the "court cipher". Note that each of the three keys listed above consists of exactly 15 letters, but why this length was chosen is not clear. Had the rule been to make the cipher messages contain only 5-letter groups, the explanation would be easy: 15 is a multiple of 5 and this would be of practical value in checking the cryptographic work. But, as has been clearly stated, disguising word lengths was apparently not the practice even if it was prescribed, so that there was no advantage in choosing keys which contain a multiple of 5 letters. And, by the way, doesn't the key COME RETRIBUTION sound rather ominous to you even these days?

Sooner or later a Confederate signal officer was bound to come up

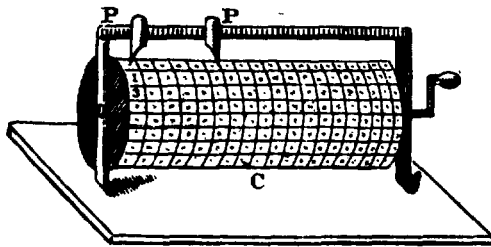
<sup>7</sup> *Battles and Leaders of the Civil War*, New York: The Century Co., 1884, Vol. I, p. 581.

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

with a device to simplify ciphering operations, and a gadget devised by a Captain William N. Barker seemed to meet the need. In Myer's *Manual* there is a picture of one form of the device, shown here in Fig. 13. I don't think it necessary to explain how it worked, for it is almost self-evident. Several of these devices were captured during the war, one of them being among the items in the NSA Museum (Fig. 14). This device was captured at Mobile in 1865. All it did was to mechanize, in a rather inefficient manner, the use of the Vigenère Cipher. But here's a photograph, Fig. 15, of the one found in the office of Confederate Secretary of State Judah P. Benjamin after the capture of Richmond. In this picture the Vigenère Square (wrapped around the revolvable central shaft) is seen very clearly.



Cipher Reel.

Fig. 13.

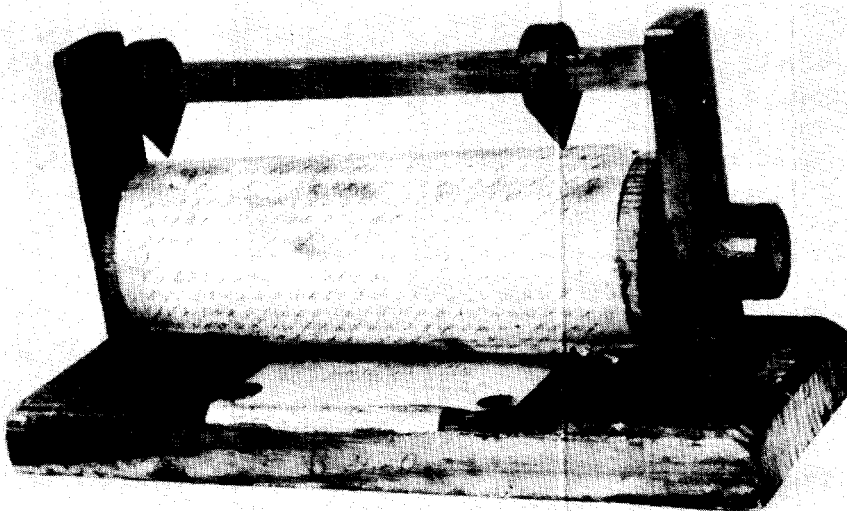


Fig. 14.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

Fig. 15.

How many of these devices were in existence or use is unknown, for their construction was an individual matter—apparently it was not an item of regular issue to members of the corps.

In practically every account of the codes and ciphers of the Civil War you will find references to ciphers used by Confederate secret service agents engaged in espionage in the North as well as in Canada. In particular, much attention is given to a set of letters in cipher, which were intercepted by the New York City Postmaster and which were involved in a plot to print Confederate currency and bonds. Much ado was made about the solution of these ciphers by cipher operators of the USMTC in Washington and the consequent breaking up of the plot. But I won't go into these ciphers for two reasons. First, the alphabets were all of the simple monoalphabetic type, a total of six altogether being used. Since they were composed of a different series of symbols for each alphabet, it was possible to compose a cipher word by jumping from one series to another without any external indication of the shift. However, good eyesight and a bit of patience were all that was required for solution in this case because of the inept manner in which the system was used: whole words, sometimes several successive words, were enciphered by the same alphabet. But the second reason for my not going into the story is that my friend and colleague of my NSA days Edwin C. Fishel, has done some research among the records in our National Archives dealing with this case and he has found something which is of great interest and which I feel bound to leave for him to tell at some future time, as that is his story, not mine.

~~CONFIDENTIAL~~

W. F. FRIEDMAN

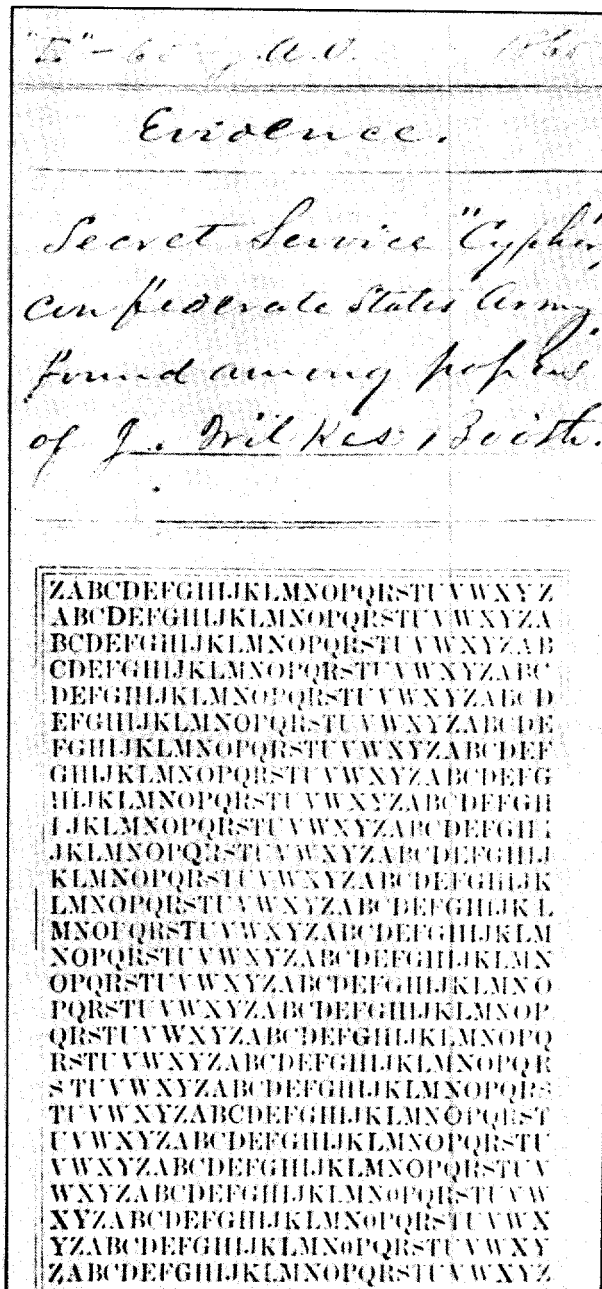
~~CONFIDENTIAL~~

Fig. 16.

Photographs from which Figs. 15 and 16 were reproduced were kindly supplied me by my friend William H. Price, of NSA.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

So very fragmentary was the amount of cryptologic information known to the general public in these days that when there was found on John Wilkes Booth's body a cipher square which was almost identical with the cipher square which had been mounted on the cipher reel found in Confederate Secretary of State Judah P. Benjamin's office in Richmond, the Federal authorities in Washington attempted to prove that this necessarily meant that the Confederate leaders were implicated in the plot to assassinate Lincoln, and had been giving Booth instructions in cipher. Fig. 16 is a picture of the cipher square found on Booth, and also in a trunk in his hotel room in Washington.

The following is quoted from Philip Van Doren Stern's book entitled *Secret Missions of the Civil War* (Rand McNally and Co., New York, 1951, p. 320):

Everyone in the War Department who was familiar with cryptography knew that the Vigenère was the customary Confederate cipher and that for a Confederate agent (which Booth is known to have been) to possess a copy of a variation of it meant no more than if a telegraph operator was captured with a copy of the Morse Code. Hundreds—and perhaps thousands of people were using the Vigenère. But the Government was desperately seeking evidence against the Confederate leaders so they took advantage of the atmosphere of mystery which has always surrounded cryptography and used it to confuse the public and the press. This shabby trick gained nothing, for the leaders of the Confederacy eventually had to be let go for lack of evidence.

To the foregoing I will comment that I doubt very much whether "everyone in the War Department who was familiar with cryptography knew that the Vigenère was the customary Confederate cipher." Probably not one of them had even heard the name Vigenère or had even seen a copy of the table, except those captured in operations. I doubt whether anyone on either side even knew that the cipher used by the Confederacy had a name; or least of all, that a German Army reservist named Kasiski, in a book published in 1863, showed how the Vigenère cipher could be solved by a straightforward mathematical method.

I have devoted a good deal more attention to the methods and means for cryptocommunications in the Civil War than they deserve, because professional cryptologists of 1961 can hardly be impressed either by their efficacy from the point of view of ease and rapidity in the cryptographic processing, or by the degree of the technical security they imparted to the messages they were intended to protect. Not much can be said for the security of the visual signaling systems used in the combat zone by the Federal Signal Corps for tactical purposes, because they were practically all based upon simple mono-

~~CONFIDENTIAL~~



W. F. FRIEDMAN

~~CONFIDENTIAL~~

alphabetic ciphers, or variations thereof, as for instance, when whole words were enciphered by the same alphabet. There is plenty of evidence that Confederate signalmen were more or less regularly reading and solving those signals. What can be said about the security of the route ciphers used by the USMTC for strategic or high command communications in the zone of the interior? It has already been indicated that, according to accounts by ex-USMTC men, such ciphers were beyond the cryptanalytic capabilities of Confederate cryptanalysts, but can we really believe that this was true? Considering the simplicity of these route ciphers and the undoubted intellectual capacities of Confederate officers and soldiers, why should messages in these systems have resisted cryptanalytic attack? In many cases the general subject matter of a message and perhaps a number of specific items of information could be detected by quick inspection of the message. Certainly, if it were not for the so-called "arbitrariness" the general sense of the message could be found by a few minutes work, since the basic system must have been known through the capture of cipher books, a fact mentioned several times in the literature. Capture of but one book (they were all generally alike) would have told Confederate signalmen exactly how the system worked and this would naturally give away the basic secret of the superseding book. So we must see that whatever degree of protection these route ciphers afforded, message security depended almost entirely upon the number of "arbitrariness" actually used in practice. A review of such messages as are available shows wide divergencies in the use of "arbitrariness". In any event the number actually present in these books must have fallen far short of the number needed to give the real protection that a well-constructed code can give. Thus it seems to me that the application of native intelligence, with some patience, should have been sufficient to solve USMTC messages—or so it would be quite logical to assume. That such an assumption is well warranted is readily demonstrable.

It was, curiously enough, at about this point in preparing this lecture that Mr. Edwin C. Fishel, whom I have mentioned before, gave me just the right material for such a demonstration. In June of 1960, Mr. Fishel had given Mr. Phillip Bridges, who is also a member of NSA and who knew nothing about the route ciphers of the USMTC, the following authentic message sent on 1 July 1863 by General George G. Meade, at Harrisburg, Pennsylvania, to General Couch at Washington. (See fig. 17.)

It took Mr. Bridges only a few hours, five or six, to solve the cryptogram, and he handed the following plain text to Mr. Fishel:

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

J. Caldwell  
 Capt Maj M. Robert  
 Nash.

Thomas for and tomorrow and acquainted  
 wish this in Chambersburg optic tree battle  
 occupy of have know a\* scouts a t. I  
 been parson get morning some with to way  
 direction I great deal soon signed and  
 concentrated rebels you are gentleman by  
 try it. Shall you reliable who country of  
 all Gettysburg Carlisle very much.

56  $\frac{117}{168}$  Pd

OWS

Fig. 17.

Thomas been it----- (Nulls)

For Parson. I shall try and get to you by tomorrow morning a reliable gentleman and some scouts who are acquainted with a country you wish to know of. Rebels this way have all concentrated in direction of Gettysburg and Chambersburg. I occupy Carlisle. Signed Optic. Great battle very soon. tree much deal---- (Nulls)

The foregoing solution is correct, save for one pardonable error: "Thomas" is not a "null" but an indicator for the dimensions of the matrix and the route. "Parson" and "Optic" are code names and I imagine that Mr. Bridges recognized them as such but, of course, he had no way of interpreting them, except perhaps by making a careful study of the events and commanders involved in the impending action, a study he wasn't called upon to undertake.

The foregoing message was enciphered by Cipher Book No. 12, in which the indicator THOMAS specifies a "Message of 10 lines and 5 columns". The route was quite simple and straightforward: "Down the 1st (column), up the 3rd; down the 2nd; up the 5th down the 4th."

It is obvious that in this example the absence of many "arbitrariness," made solution a relatively easy matter. What Mr. Bridges would have been able to do with the cryptogram had there been many of them is problematical. Judging by his worksheets, it seemed to me

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

that Mr. Bridges did not realize when he was solving the message that a transposition matrix was involved; and on questioning him on this point his answer was in the negative. He realized this only later.

A minor drama in the fortunes of Major General D. C. Buell, one of the high commanders of the Federal Army, is quietly and tersely outlined in two cipher telegrams. The first one, sent on 29 September 1862, from Louisville, Kentucky, was in one of the USMTC cipher books, and was externally addressed to Colonel Anson Stager, head of the USMTC, but the internal addressee was Major General H. W. Halleck, "General-in-Chief" [our present day "Chief of Staff"]. The message was externally signed by William H. Drake, Buell's cipher operator, but the name of the actual sender, Buell, was indicated internally. Here's the telegram:

COLONEL ANSON STAGER, Washington:

Austria await I in over to requiring orders olden rapture blissful for your instant command turned and instructions and rough looking further shall further the Camden me of ocean September poker twenty I the to I command obedience repair orders quickly pretty Indianapolis your him accordingly my fourth received 1862 wounded nine have twenty turn have to to to alvord hasty.

WILLIAM H. DRAKE

Rather than give you the plain text of this message, perhaps you would like to work it out for yourselves, for with the information you've already received the solution should not be difficult. The message contains one error, which was made in its original preparation: one word was omitted.

The second telegram, only one day later, was also from Major General Buell, to Major General Halleck, but it was in another cipher book—apparently the two books involved were used concurrently. Here it is:

GEORGE C. MAYNARD, Washington:

Regulars ordered of my to public out suspending received 1862 spoiled thirty I dispatch command of continue of best otherwise worst Arabia my command discharge duty of my last for Lincoln September period your from sense shall duties the until Seward ability to the I a removal evening Adam herald tribune.<sup>8</sup>

PHILIP BRUNER

As before, I will give you the opportunity to solve this message

<sup>8</sup> A curious coincidence—or was it a fortuitous foreshadowing of an event far in the future?—can be seen in the sequence of the last two words of the cipher text. The message is dated September 30, 1862; the New York Herald and the New York Tribune combined to make the New York Herald-Tribune on March 19, 1924—62 years later!

**Fig. 18-A.**

W. F. FRIEDMAN

~~CONFIDENTIAL~~

Greensboro N.C.

April 11 1865

Benaja 11 Hd Q near R. G.

Genl J. E. Johnston

A scout (reports?) that Genl Lee

u i D v v s w v z F x - m q s - E G A z o x -

H W - P J M - T z A t - near to appomattox Court

house yesterday No official intelligence of the

event D i F - x y i k v - q T - F B B H Y G -

F A S D - J H i - L P O u B - As to result Gen H. H.

Walker is ordered Y W F T - W S K T M T - B X z S -

G q - X A m E - C H T - i u - A K M S A u P u V F -

Let me hear from you there- I will have need to

see you to confer as to future action. The above

is my telegram of yesterday which is repeated as

requested.

Jeffn Davis

Official

Burton Harrison

Private Secty

18-B.

65

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

for yourselves. (At the end of the next lecture I shall present the plain text of both messages.)

Figure 18 is a photograph of an important message which you may wish to solve yourself. It was sent by President Jefferson Davis to General Johnston, on a very significant date, 11 April 1865.\* For ease in working on it I give also a transcription, since the photograph is very old and in poor state. I believe that this message does not appear in any of the accounts I've read.

It is time now to tell you what I can about the success or lack of success which each side had with the cryptograms of the other side. I wish there were more information on this interesting subject than what I am about to present. Most of what sound information there is comes from a book by a man named J. Willard Brown, who served four full years in the Federal Army's Signal Corps. The book is entitled *The Signal Corps, U.S.A. in the War of the Rebellion*, published in Boston in 1896 by the U.S. Veteran Signal Corps Association. In his book Brown deals with the cryptanalytic success of both sides. First, let's see what the Union signalmen could do with rebel ciphers. Here are some statements he makes (p. 214):

The first deciphering of a rebel signal code of which I find any record was that made by Capt. J. S. Hall and Capt. R. A. Taylor, reported Nov. 25, 1862. Four days later, Maj. Myer wrote to Capt. Cushing, Chief Signal Officer, Army of the Potomac, not to permit it to become public "that we translate the signal messages of the rebel army".

April 9, 1863, Capt. Fisher, near Falmouth, reported that one of his officers had read a rebel message which proved that the rebels were in possession of our code. The next day he was informed that the rebel code taken (from) a rebel signal officer was identical with one taken previously at Yorktown.

He received from Maj. Myer the following orders:

"Send over your lines, from time to time, messages which, if it is in the power of the enemy to decipher them, will lead them to believe that we cannot get any clew to their signals."

"Send also occasionally messages untrue, in reference to imaginary military movements, as for instance,—“The Sixth Corps is ordered to reinforce Keyes at Yorktown.”"

Undoubtedly, what we have here are references to the general cipher system used by the Confederates in their electric-telegraph communications, for note the expression "Send over your lines". This could hardly refer to visual communications. Here we also have very early instances, in telegraphic communications, of what we call cover and deception, i. e., employing certain ruses to try to hide the fact that enemy signals could be read, and to try to deceive him

---

\*I should warn you that it contains several errors!

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

by sending spurious messages for him to read, hoping the fraud will not be detected.

Brown's account of Union cryptanalytic successes continues (p. 215):

In October, 1863, Capt. Merrill's party deciphered a code, and in November of the same year Capt. Thickstun and Capt. Marston deciphered another in Virginia.

Lieut. Howgate and Lieut. Flook, in March, 1864, deciphered a code in the Western Army, and at the same time Lieut. Benner found one at Alexandria, Virginia.

Capt. Paul Babcock, Jr., then Chief Signal Officer, Department of the Cumberland, in a letter dated Chattanooga, Tennessee, April 26, 1864, transmitting a copy of the rebel signal code, says:

Capt. Cole and Lieut. Howgate, acting Signal Officers, occupy a station of communication and observation on White Oak Ridge at Ringgold, Ga. . . . On the 22nd inst. the rebels changed their code to the one enclosed, and on the same day the above-mentioned officers by untiring zeal and energy succeeded in translating the new code, and these officers have been ever since reading every message sent over the rebel lines. Many of these messages have furnished valuable information to the general commanding the department.

The following is also from Brown (p. 279):

About the first of June (1864), Sergt. Colvin was stationed at Fort Strong, on Morris Island, with the several codes heretofore used by the rebels, for the purpose of reading the enemy's signals if possible. For nearly two weeks nothing could be made out of their signals, but by persevering he finally succeeded in learning their codes. Messages were read by him from Beach Inlet, Battery Bee, and Fort Johnson. Gen. J. G. Foster, who had assumed command of the Department of the South, May 26th, was so much pleased with Sergt. Colvin's work, that in a letter addressed to Gen. Halleck, he recommended "that he be rewarded by promotion to Lieutenant in the Signal Corps, or by a brevet or medal of honor." This recommendation was subsequently acted upon, but, through congressional and official wrangling over appointments in the Corps, he was not commissioned until May 13, 1865, his commission dating from Feb. 14, 1865.

(p. 281):

During the month, Sergt. Colvin added additional laurels to the fame he had earned as a successful interpreter of rebel signals. The enemy had adopted a new cipher for the transmission of important messages, and the labor of deciphering it devolved upon the sergeant. Continued watchfulness at last secured the desired result, and he was again able to translate the important dispatches of the enemy for the benefit of our commandants. The information thus gained was frequently of special value in our operations, and the peculiar ability exhibited by the sergeant led Gen. Foster once more to recommend his promotion.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

(p. 286):

About the same time an expedition under Gen. Potter was organized to act in conjunction with the navy in the vicinity of Bull's Bay. Lieut. Fisher was with this command, and by maintaining communications between the land and naval forces facilitated greatly the conjoined action of the command. Meanwhile every means was employed to intercept rebel messages. Sergt. Colvin, assigned to this particular duty, read all the messages within sight, and when the evacuation of Charleston was determined upon by the enemy, the first notification of the fact came in this way before the retreat had actually commenced. As a reward for conspicuous services rendered in this capacity, Capt. Merrill recommended that the sergeant be allowed a medal, his zeal, energy and labors fully warranting the honor.

After the occupation of Charleston, communications was established by signals with Fort Strong, on Morris Island, Fort Johnson and James Island, Mount Pleasant, and Steynmeyer's Mills. A line was also opened with the position occupied by the troops on the south side of the Ashley river.

With regard to Confederate reading of Union visual signals, Brown makes the following observations of considerable interest (p. 274):

The absolute necessity of using a cipher when signalling in the presence of the enemy was demonstrated during these autumn months by the ease with which the rebels read our messages. This led to the issuing of an order that all important messages should be sent in cipher. Among the multitude of messages intercepted by the enemy, the following were some of the more important . . .

Brown thereupon cites 25 such messages but he gives no indication whatever as to the source from which he obtained these examples or how he knew they had been intercepted. They all appear to be tactical messages sent by visual signals.

In many of the cases cited by Brown it is difficult to tell whether wig-wag or electric telegraph messages were involved. But in one case, (evacuation of Charleston) it is perfectly clear that visual messages were involved, when Brown says that Sgt. Colvin "read all the messages within sight."

Further with regard to rebel cryptanalytic success with Union messages, Brown has this to say (p. 213):

The reports of Lieut. Frank Markoe, Signal Officer at Charleston, show that during the siege thousands of messages were sent from one post to another, and from outposts to headquarters, most of which could have been sent in no other way, and many were of great importance to the Confederate authorities.

Lieut. Markoe says that he read nearly every message we sent. He was forewarned of our attack on the 18th of July, 1863. He adds regretfully, however, that through carelessness of the staff officers at headquarters it leaked out that he was reading our messages. Our officers then began to use the cipher disk. In August he intercepted

~~CONFIDENTIAL~~



W. F. FRIEDMAN

~~CONFIDENTIAL~~

the following message: "Send me a copy of rebel code immediately, if you have one in your possession". He therefore changed his code. . . . A little later our officers used a cipher which Lieut. Markoe says he was utterly unable to unravel.

It is unfortunate that neither Lieutenant Markoe, the Confederate cryptanalyst, nor Brown, the Union signalman, tell us what sort of cipher this was that couldn't be unravelled. I assume that it was the Myer disk used properly, with a key phrase of some length and with successive letters, not whole words, being enciphered by successive letters of the key. But this is only an assumption and may be entirely erroneous.

In the foregoing citations of cryptanalytic successes it is significant to note that visual messages were intercepted and read by both sides; second, that Confederate telegraphic messages protected by the Vigenère cipher were read by Union personnel whenever such messages were intercepted; and third, that USMTC telegraph messages protected by the route cipher, apparently intercepted occasionally, were never solved. Later I shall make some comments on this last statement, but at the moment let us note that technically the Vigenère cipher is theoretically much stronger than the route cipher, so that we have here an interesting situation, *viz*, the users of a technically inferior cryptosystem were able to read enemy messages protected by a technically superior one, but the users of a technically superior cryptosystem were not able to read enemy messages protected by a technically inferior one—a curious situation indeed.

I can hardly close this lecture without citing a couple of messages which appear in nearly every account I've seen of the codes and ciphers of the Civil War. These are messages which were sent by President Lincoln under circumstances in which, allegedly, the usual cipher could not be or, at least was not, employed. The first of the two was sent on 25 November 1862 from the White House to Major General Burnside, Falmouth, Virginia. The circumstances are so bizarre that if I merely presented the cipher message to you without some background I doubt if you would believe me. And even after I've presented the background, I'm sure you won't know what to think. I, myself, don't really know whether to take the incident seriously or not. Let me quote from an account of it in the book by David Homer Bates, one of the first members of the USMTC, in his *Lincoln in the Telegraph Office* (D. Appleton-Century Co., New York, 1939, pp. 58-61):

"During Burnside's Fredericksburg campaign at the end of 1862, the War Department operators discovered indications of an interloper on the wire leading to his headquarters at Aquia Creek. These indications consisted of an occasional irregular opening and closing

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

of the circuit and once in a while strange signals, evidently not made by our own operators. It is proper to note that the characteristics of each Morse operator's sending are just as pronounced and as easily recognized as those of ordinary handwriting, so that when a message is transmitted over a wire, the identity of the sender may readily be known to any other operator within hearing who has ever worked with him. A somewhat similar means of personal identification occurs every day in the use of the telephone.

"At the time referred to, therefore, we were certain that our wire had been tapped. In some way or other the Confederate operator learned that we were aware of his presence, and he then informed us that he was from Lee's army and had been on our wire for several days, and that, having learned all that he wanted to know, he was then about to cut out and run. We gossiped with him for a while and then ceased to hear his signals and believed that he had gone.

"We had taken measures, however, to discover his whereabouts by sending out linemen to patrol the line; but his tracks were well concealed, and it was only after the intruder had left that we found the place where our wire had been tapped. He had made the secret connection by means of fine silk-covered magnet wire, in such a manner as to conceal the joint almost entirely. Meantime, Burnside's cipher-operator was temporarily absent from his post, and we had recourse to a crude plan for concealing the text of telegrams to the Army of the Potomac, which we had followed on other somewhat similar occasions when we believed the addressee or operator at the distant point (not provided with the cipher-key) was particularly keen and alert. This plan consisted primarily of sending the message backward, the individual words being misspelled and otherwise garbled. We had practised on one or two despatches to Burnside before the Confederate operator was discovered to be on the wire, and were pleased to get his prompt answers, couched also in similar outlandish language, which was, however, intelligible to us after a short study of the text in each case. Burnside and ourselves soon became quite expert in this home-made cipher game, as we all strove hard to clothe the despatches in strange, uncouth garb.

"In order to deceive the Confederate operator, however, we sent to Burnside a number of cipher messages, easy of translation, and which contained all sorts of bogus information for the purpose of misleading the enemy. Burnside or his operator at once surmised our purpose, and the general thereupon sent us in reply a lot of balderdash also calculated to deceive the uninitiated.

"It was about this time that the following specially important despatch from Lincoln was filed for transmission:

Executive Mansion, Washington,  
November 25, 1862. 11:30 AM.

MAJOR-GENERAL BURNSIDE, Falmouth, Virginia: If I should be in boat off Aquia Creek at dark to-morrow (Wednesday) evening, could you, without inconvenience, meet me and pass an hour or two with me?

A. Lincoln.

"Although the Confederate operator had said good-by several days

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

before, we were not sure he had actually left. We therefore put Lincoln's telegram in our home-made cipher, so that if the foreign operator were still on our wire, the message might not be readily made out by the enemy. At the same time extra precautions were taken by the Washington authorities to guard against any accident to the President while on his visit to Burnside. No record is now found of the actual text of this cipher-despatch, as finally prepared for transmission, but going back over it word for word, I believe the following is so nearly like it as to be called a true copy:

Washington, D. C., November 25, 1862

BURNSIDE, Falmouth, Virginia: Can Inn Ale me withe 2 oar our Ann pas Ann me flesh ends N. V. Corn Inn out with U cud Inn heaven day nest Wed roe Moore Tom darkey hat Greek Why Hawk of Abbott Inn B chewed I if. BATES.

This sort of subterfuge is hardly worthy of becoming embalmed in the official records of the war—and apparently it wasn't. But several years later, one of identical nature did become so embalmed, for the message appears on page 236, Vol. 45, of "Telegrams received by the Secretary of War":

Hq. Armies of the U. S., City Point, Va.,  
8:30 a. m., April 3, 1865

TINKER, War Department: A. Lincoln its in fume a in hymn to start I army treating there possible if of cut too forward pushing is He is so all Richmond aunt confide is Andy evacuated Petersburg reports Grant morning this Washington Secretary War. BECKWITH.

Both Plum and Bates cite the foregoing telegram and their comments are interesting if not very illuminating. Plum says merely: "By reading the above backward with regard to the phonetics rather than the orthography, the meaning will be apparent". Bates says:

"The probable reason for adopting this crude form was to insure its reaching its destination without attracting the special attention of watchful operators on the route of the City Point-Washington wire, because at that crisis every one was on the *Qui vive* for news from Grant's advancing army, and if the message had been sent in plain language, the important information it conveyed might have been overheard in its transmission and perhaps would have reached the general public in advance of its receipt by the War Department.

"It is not necessary to give the translation of this cipher-message. To use a homely term, 'Any one can read it with his eyes shut.' In fact, the easiest way would be for one to shut the eyes and let some one else read it backward, not too slowly. The real wording then becomes plain.

Can you imagine for one moment that a "cryptogram" of such simplicity could not be read at sight by any USMTC operator, even without having someone read it to him backward? Such a "cryptogram" is hardly worthy of a schoolboy's initial effort at preparing a

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

secret message. But I assure you that I did not make this story up, nor did I compose the cryptogram.

Ruminating upon what I have shown and told you about the cryptosystems used by both sides in the Civil War, do you get the feeling, as I do, that the cryptologic achievements of neither side can be said to add lustre to undoubtedly great accomplishments on the battlefield? Perhaps this is a good place to make an appraisal of the cryptologic efficiency of each side.

First, it is fair to say that we can hardly be impressed with the cryptosystems used by either side. The respective Signal Corps at first transmitted by visual signals messages wholly in plain language; such messages were often intercepted and read straight-away. Then both sides began enciphering such messages, the Signal Corps of the Federal Army using a cipher disk invented by the Chief Signal Officer, the Signal Corps of the Confederate Army using the Vigenère cipher. In both cases the use of cryptography for tactical messages was quite inept, although it seems that from time to time the Federal signalmen had better success with the Vigenère-enciphered visual messages of the Confederate signalmen than the latter had with the disk-enciphered messages of the Union signalmen.

With regard to the cryptosystem used by the Confederate Signal Corps, although there may initially have been cases in which mono-alphabetic substitution alphabets were used, such alphabets were probably drawn up by agreement with the signal officers concerned, and changed from time to time. Nowhere have I come across a statement that the Myer disk or something similar was used. In any event, messages transmitted by visual signals were read from time to time by Union signalmen, the record showing a number of cases in which the latter "worked out the rebel signal code"—meaning, of course, that the substitution alphabet involved was solved. When did the Confederate Signal Corps begin using the Vigenère cipher? The answer seems to be quite clear. In a letter dated 6 June 1888 from General J. H. Alexander (brother of General E. P.) to J. Willard Brown<sup>9</sup> we find the following statements:

"At the first inauguration of the Signal Service in the Confederacy, I, having received in the first place the primary instruction from my brother, Gen. E. P. A., then a colonel on Beauregard's staff near the Stone Bridge at Manassas, was assigned the duty of preparing a confidential circular of instruction for the initiation of officers and men, in this branch. I did prepare it, in Richmond, in early spring, 1862, and surrendered the copy to Hon. James A. Seddon, the then Secretary of War at Richmond. It was issued in form of a small pamphlet. I had attached a table for compiling cipher dispatches—which was printed

<sup>9</sup> Op. Cit., p. 206.

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

*with the rest of the matter—and the whole was issued confidentially to the officers newly appointed for signal duty. (My emphasis)*

I have italicized the last sentence because I think that the "table for compiling cipher dispatches" can refer only to the Vigenère square table, for that and only that sort of table is even mentioned in accounts of the ciphers used by the Confederacy. One could, of course, wish that the writer had given some further details but there are none. However, the statement about the table is sufficiently explicit to warrant the belief that it was General J. H. Alexander who officially introduced the Vigenère square into Confederate cryptography, although he may have obtained the idea from his brother, since he states that he "received in the first place the primary instruction from my brother".

In the Federal Signal Corps it is quite possible that the polyalphabetic methods Myer cites in his *Manual* for using his cipher disk (changing the setting with successive words of a message) were used in some cases, because there are found in the record several instances in which the Confederate signalmen, successful with monoalphabetic encipherments, were completely baffled. One is warranted in the belief that it was not so much the complexities introduced by using a keyword to encipher successive *words* of the plain text as it was the lack of training and experience in cryptanalysis which hampered Confederate signalmen who tried to solve such messages. In World War I a German Army system of somewhat similar nature was regularly solved by Allied cryptanalysts, but it must be remembered, in the first place, that by 1914 the use of radio made it possible to intercept volumes of traffic entirely impossible to obtain before the advent of radiotelegraphy; and, in the second place, would-be cryptanalysts of both sides in the Civil War had nothing but native wit and intelligence to guide them in their work on intercepted messages, for there were, so far as the record goes, no training courses in *cryptanalysis* on either side, though there were courses in cryptography and signaling. It would seem to cryptanalysts of 1961, a century later, that native wit and intelligence nevertheless should have been sufficient to solve practically every message intercepted by either side, so simple and inefficient in usage do the cryptosystems employed by both sides appear today.

No system employed by the Federals, either for tactical messages (Signal Corps transmissions) or strategic messages (USMTC transmissions) would long resist solution today, provided, of course, that a modicum of traffic were available for study. Although technically far less secure in actual practice than properly enciphered Vigenère messages, the route ciphers of the USMTC seem to have eluded the efforts of inept Confederate cryptanalysts. Ex-USMTC operators

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

make the statement that none of their messages was ever solved and that the Confederates published intercepted messages in Southern newspapers in the hope that somebody would come forward with a solution; yet it must be remembered that those operators were Northerners who were very naturally interested in making the achievements of the Union operators, both in cryptography and in cryptanalysis, appear more spectacular than they really were. And it is probable that they wrote without having made a real effort to ascertain whether the Confederates did have any success. A "real effort" would have been a rather imposing undertaking then—as it still is, I fear. Now it must be presumed that if Confederate operators had succeeded in solving intercepted traffic of the USMTC they would have recorded the facts to their own credit. But in his seven volumes on the campaigns of Lee and his lieutenants, Douglas S. Freeman does not mention a single instance of interception and solution of telegraphic messages of the Union. Perhaps Freeman was seeking 100% confirmation, which is too much to expect in a field of such great secrecy. This failure of the Confederate cryptanalysts is the more astonishing when we know that copies of the USMTC cipher books were captured and that, therefore, they must have become aware of the nature of the route ciphers used by the USMTC, unless there was a lack of appreciation of the value of such captures and a failure to forward the books to the proper authorities, who could hand them over to their experts. In those books the USMTC route ciphers would have been seen in their naive simplicity, complicated only by the use of "arbitraries" or code equivalents, but hardly to the degree where all messages would be impossible to solve. It seems to me that there can be only four possible explanations for this failure to solve the USMTC route ciphers. Let us examine them in turn.

First, it is possible that there was not enough intercept traffic to permit solution. But this is inadequate as an explanation. The route cipher is of such simplicity that "depth" is hardly an absolute requirement—a single message can be solved, and its intelligibility will be determined to a large degree by the number of "arbitraries" it contains. Where there are many, only the dim outlines of what is being conveyed by the message may become visible; where there are few or even none, the meaning of the messages becomes fairly evident. But the abundant records, although they contain many references to intercepts, fail to disclose even one instance of solution of a USMTC message. Thus we are forced to conclude that it was not the lack of intercept traffic which accounts for lack of success by the Confederates with USMTC messages, but some other factor.

~~CONFIDENTIAL~~

W. F. FRIEDMAN

~~CONFIDENTIAL~~

Second, the lack of training in cryptanalysis of Confederate cryptanalysts might have been the reason why Confederate signalmen failed to solve the messages. This sounds plausible until we look into the matter with a critical spirit. Solution of route ciphers requires little training; native wit and intelligence should have been sufficient. The degree of intelligence possessed by Confederate officers and men was certainly as high as that of their Union counterparts who were up against a technically far superior cryptosystem, the Vigenère. We may safely conclude that it was not lack of native wit and intelligence that prevented them from solving messages enciphered by the USMTC route ciphers.

Third, it is possible that Confederate high commanders were not interested in communications-intelligence operations or in gathering the fruits of such operations. Such an explanation seems on its face fatuous and wholly unacceptable. We know of the high estimate of value field commanders placed upon the interception and solution of tactical messages transmitted by visual signaling; but an appreciation of the extraordinary advantages of learning the contents of enemy communications on the strategic level may have been lacking. My colleague Mr. Fishel thinks that "intelligence consciousness" and "intelligence sophistication" were of a very low order in the Union Army, and of a markedly lower order in the Confederate Army. But to us, in 1961, to disregard the advantages of a possible reading of strategic messages seems almost incredible and I am inclined to discount this sort of explanation.

Fourth, it is possible that Confederate cryptanalysts were far more successful in their efforts to solve USMTC transmissions than present publicly-available records indicate; that Confederate commanders obtained great advantages from their communications-intelligence operations; that they fully recognized the supreme necessity of keeping this fact and these advantages secret; and that the Confederate States Government adopted and enforced strict communications-intelligence security regulations, so that the truth concerning these matters has not yet emerged. Let it be noted in this connection that very little information can be found in the public domain today about Allied cryptanalytic successes during World War I; and were it not for the very intensive and extensive investigations in the matter of the Japanese attack on Pearl Harbor on 7 December 1941, very little, if any, information would be known to the public about British and American successes in communications-intelligence during World War II. Immediately following the capture of Richmond and before Confederate records could be removed to a safe place, a great fire broke out and practically all those records were destroyed. It is possible that this is one of the reasons why the

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

records of their communications-intelligence successes have never come to light. But it is also possible that Confederate cryptanalysts kept their secrets to themselves. We know that the records possessed or taken by certain Confederate leaders have been gone over with great care and attention, but what happened to those retained by other Confederate leaders such as the Secretary of War Seddon, or his predecessor Judah P. Benjamin, who later became Secretary of State, and others? Here is a fascinating speculation and one which might well repay careful, painstaking research in the voluminous records of our National Archives. I shall leave the delving into those records to some of you young and aspiring professional cryptanalysts who may be interested in undertaking such a piece of research. With this thought I bring this lecture to its close.

~~CONFIDENTIAL~~



UNCLASSIFIED

## The Association Factor in Information Retrieval

BY H. EDMUND STILES

*Unclassified*

*This paper describes an all-computer document-retrieval system which can find documents related to a request even though they may not be indexed by the exact terms of the request, and can present these documents in the order of their relevance to the request.*

All documentalists who are operating large coordinate indexes are searching for better ways to exploit this type of information system. In our library we have already eliminated the time-consuming job of posting document numbers manually by enlisting the aid of a 705 computer. (The computer periodically prepares revised posting cards to replace the outdated ones.) Now we are searching for better solutions to our retrieval problems.

One obvious retrieval problem in any large system is the time required to "coordinate" heavily posted terms. We are convinced we must mechanize if we are to allow our collection to grow indefinitely.

A second problem is the retrieval of so many documents related to a single request that the customer finds it difficult to decide which document to examine first. Since he has no precise means of determining which document is most closely related to a request, we have tried to assist him in using somewhat arbitrary or subjective means. The date of the document is sometimes used as a relevance criterion, in the hope that the most recent document will be the most pertinent, or the name of the author is used, in the hope that the work of a known author will answer the request better than that of an unknown one. The pitfalls of such criteria are apparent.

The third, and by far the most serious difficulty in a large system, is the problem of choosing terms for search which will turn up all of the documents relevant to the request. Our handicap has been that we have had to select the precise terms that were originally used to index the desired document. Literally hundreds of terms may have been used to index documents on the various aspects of a particular subject and yet we must grope for just the right set of terms. Just as the indexer tried to use language which he hoped would be used by future requesters, so the requester must hope to use the same

---

This article has been published in the April 1961 issue of *The Journal of the ACM* (Association for Computing Machinery).

UNCLASSIFIED

## UNCLASSIFIED INFORMATION RETRIEVAL

terms that were used by the indexer in processing the required documents.

With our new method we believe we can overcome all three difficulties. First, every step of the process can be performed by existing machines; second, in answer to a given request our machines will deliver a list of documents arranged in the approximate order of their relevance to the request, and third, we will be able to find these documents even though they may not be indexed by the terms of the initial request.

Our general strategy is to generate by machine an expanded list of request terms that will serve as a bigger net to catch documents. Once caught we will grade them automatically so that the most important ones will be on top. Our experiment was conducted on an existing collection of over 100,000 documents already indexed by the Uniterm Coordinate Index System. [1]

The first step in our procedure is to develop a list of terms arranged according to their degree of association with a given term. Frequency alone is not a satisfactory measure of association. For example, we counted the number of times various terms had been used together with the term "Friction" to index a document and found that of the 105 terms used, the most frequent were:

Theory.....	7 times
Film.....	6
Crystal.....	5
Metal.....	5
Thin.....	5
Transfer.....	4
Clutch.....	3
Damping.....	3
Electrostatic.....	3

Although "Metal" and "Clutch" may be significantly associated with "Friction", obviously the word "Theory" which is at the top of the list has no more relationship to "Friction" than to any other word about which there might be a theory. We searched for a formula that would give us a relative frequency—one that would measure the distance from the expected frequency of occurrence assuming no association. After considering several other formulas, including the ones reported by Maron, Kuhns and Ray in their report on "Probabilistic Indexing", [2] we decided to use the following:

$$\log_{10} \frac{\left( \left| fN - AB \right| - \frac{N}{2} \right)^2 N}{AB (N - A) (N - B)} = \text{ASSOCIATION FACTOR},$$

H. E. STILES

UNCLASSIFIED

where  $A$  is the number of documents indexed by one term;

$B$  is the number of documents indexed by a second term;

$f$  is the number of documents indexed by the combination of both terms; and

$N$  is the total number of documents in the collection.

This formula is a form of the chi square formula using the marginal values of the  $2 \times 2$  contingency table and the Yates' correction [3] for small samples. If  $AB > fN$  the association is negative. Such occurrences must be recognized during the computation process and the resultant *association factors* marked to indicate negative association. By applying this formula to each of the 105 terms paired with "Friction" the top of the list became as follows:

<i>Term</i>	<i>f</i>	<i>A</i>	<i>B</i>	<i>Association Factor</i>
Wear	2	4	25	3.35
Thin	5	49	25	3.21
Lubrication	2	9	25	3.00
Belt	1	2	25	2.70

"Theory" dropped down to a much more reasonable position, and terms such as "Analysis", "Problems" and "Study" were at the bottom. "Wear" had risen to the top even though it occurred only twice in association with "Friction". Anyone interested in friction would probably be interested in the two additional documents indexed by "Wear" and the seven additional documents indexed by "Lubrication".

We tried the same experiment for the term "Exposure" with the following results:

<i>"Exposure"</i>				
<i>Term</i>	<i>f</i>	<i>A</i>	<i>B</i>	<i>Association Factor</i>
Weathering	3	3	29	3.86
Plywood	1	1	29	2.94
Nylon	2	12	29	2.80
Enamel	1	2	29	2.63
Microfilm	3	52	29	2.61
Preservatives	1	3	29	2.46
Lenses	3	77	29	2.44
Radiography	1	4	29	2.33
Protective	1	12	29	1.85

Terms that had association factors of less than one (1.00) were discarded. On this basis only a small portion of the terms that had been used with "Exposure" (or "Friction") were considered to be associated with it.

## UNCLASSIFIED INFORMATION RETRIEVAL

These term profiles, as we have chosen to call these lists of associated terms, have four important characteristics. First, they are derived from the document collection itself rather than from the subjective realm of human experience. Therefore only the terms that will be useful in finding documents are included and extraneous terms are eliminated. Second, they are generated in a statistical manner which can be duplicated by an unthinking computer, an encouraging fact considering the future masses of literature to be indexed. Third, they reveal the various facets of meaning that the term has in our particular collection. The profile for the term "Exposure", for instance, contains terms used when describing "exposure to the elements," "exposure of photographic film," and "exposure to radiation." This characteristic makes explicit the variety of meanings that were inherent in the parent term—a fact we will come to appreciate when we start combining them. And finally, the profiles derived by this method alone contain terms that are only statistically related and not semantically related to the request term. This distinction has been well explained by Maron, Kuhns and Ray. [2]

"Whereas the semantical relationships are based solely on the meanings of the terms and hence independent of the 'facts' described by those words, the statistical relationships between terms are based solely on the relative frequency with which they appear and hence *are* based on the nature of the facts described by the documents. Thus, although there is nothing about the meaning of the term 'logic' which implies 'switching theory', the nature of the facts (viz., that truth-functional logic is widely used for the analysis and synthesis of switching circuits) 'causes' a statistical relationship. Another example might concern the terms 'information theory' and 'Shannon' . . ."

Later we will describe how to derive semantic relationships as well as purely statistical ones.

When we have prepared a term profile for each request term we are ready to proceed to the second step, which is to compare the profiles of each term of a multiterm request and select those terms which appear in all or in a given number of profiles. These selected terms are called *first generation terms*. We are aware of the possibilities in a conventional coordinate indexing system of requesting documents that have a logical product, sum, or negation of the request terms. The same flexibility exists when using the association factor. If the request is for documents on "*American Tractor Tires*" we would prepare a vocabulary profile for each term and then select only those terms which appear in all three profiles. However, if we were interested in "*American Tractor OR Automobile Tires*" we would select those terms that appeared in the profiles of "American" and "Tires" and either "Tractor" or "Automobile". These first genera-

H. E. STILES

UNCLASSIFIED

tion terms therefore tend to reflect the logic of the request. However we cannot exclude from our first generation terms all the terms in the profile of a "not" term because of the danger of also eliminating some desirable terms. "Not" terms must be used by themselves to eliminate documents that have been indexed by them. If a request has only a single term, the terms of its profile are the same as its first generation terms.

The end result of this second step described above, is a list of first generation terms which have been used with the original request terms to index documents much more frequently than would be expected of terms having no association. Remember also, that these first generation terms are only statistically associated with the request terms. Synonyms or near synonyms are not likely to be found in this list, because documents are not usually indexed by synonymous terms. Yet synonyms, near synonyms, generics, specifics, and other closely related words would be desirable additions to an expanded list of request terms. Our method of generating these constitutes our third step. It projects us beyond the purely statistical relationships and into the realm of meaningful associations. This step is to treat the first generation terms as request terms and repeat steps 1 and 2. Since there may be quite a number of first generation terms, we need not require that a term appear in all of their profiles, but only in approximately one fifth or in some other specified number of profiles. The resultant new terms are called *second generation terms*. Among these we find words closely related in meaning to the request terms.

For example, if we were asked for all documents on United States wheat exports, the profile of the term "United States" would probably not contain the terms "Uncle Sam" or "USA" even if they were permissible in our term dictionary, since any given document would not be indexed by more than one of the three terms. Since they are missing from the "United States" profile, they would not be included among the first generation terms. However, when the first generation terms such as "Kansas", "Bushels", "Dollars", "Grain", "Cordell Hull", "Tariff", etc., are treated as request terms, each may well have "Uncle Sam" and "USA", as well as "United States" among its profiles. Assuming that they will appear in a sufficient number of the profiles, they will qualify as second generation terms.

In our coordinate index we have tried to eliminate all synonyms by cross-referencing them to a single term in our term dictionary. However, when we requested documents on the "Weatherproofing of Fabrics" we derived "Fungus", "Plastic", "Exposure", and "Coating" among the first generation terms, and "Weathering", "Fungi-

## UNCLASSIFIED INFORMATION RETRIEVAL

cidal", and "Preservatives" among the second generation terms. In future applications of this system we can expect the second generation terms to include not only synonyms, but also various grammatical forms and even variations in spelling of the request terms.

We now have an expanded list of request terms. It includes the original request terms, the first generation terms and the second generation terms. It is reasonable to assume that these terms do not all have the same degree of association with the original request and that it would be helpful to determine the degree of association for each before proceeding further.

The fourth step is the preparation of a table of the expanded list, in which we would record the association factors of each term to all others. We record only those above the established threshold of 1. The sum of the association factors for each term, divided by the total number of terms in the expanded list, gives us a weight which will enable us to arrange the terms according to their probable relevance to the request. The expanded list of terms related to the "Weatherproofing of Fabrics" with term weights is as follows:

Fabrics.....	2.67	Deterioration.....	1.26
Plastics.....	2.58	Resistance.....	1.25
Coating.....	2.38	Protection.....	1.25
Fungus.....	2.28	Agar.....	1.23
Weatherproofing.....	2.04	Metals.....	1.21
Tests.....	1.90	Plate.....	1.17
Exposure.....	1.87	Biphenyl.....	1.14
Compounds.....	1.76	Dinitrofluorotoluene.....	1.14
Laminates.....	1.73	Dinitrobenzene.....	1.14
Resins.....	1.56	Vinyl.....	1.14
Weathering.....	1.52	Preservatives.....	1.09
Materials.....	1.43	Elastomers.....	1.08
Glass.....	1.42	Molded.....	1.06
Cotton.....	1.36	Aluminum.....	1.04
Chemical.....	1.34	Aging.....	1.03
Fungicidal.....	1.32	Temperature.....	1.02
Compressor.....	1.27	Fluorine.....	1.00

No one word in this list could be substituted for the request, because each has its own variety of meanings and uses, yet it would be hard to use a group of them without touching on the subject of the request. We now have a powerful tool with which to search for documents, for we are not dependent upon the requester and the indexer using the same language. Rather, we have fashioned a request language from the consensus of all previous indexing.

We are now ready for step 5. We compare the expanded list of

H. E. STILES

UNCLASSIFIED

request terms with the index terms of each document in the collection. Whenever the terms match, the weight of the request term is assigned to the corresponding document index term. The sum of these weights for each document is called the *document relevance number*. This number should indicate the degree of fit between the request and the contents of the document.

From a request for all documents on the subject of "Thin Films" we found the list of document numbers indicated in column 1 of Table 1. These document numbers are arranged according to their *document relevance numbers* which appear in column 2. We then asked a qualified engineer to examine these documents and specify which were related to "Thin Films" and which were not. He developed his own rating scale which was as follows:

- Yes—Contains information on "Thin Film".
- M —May be useful background information.
- P —Possibly contains useful background information.
- No —Does not contain information on "Thin Film".

This engineer was not familiar with our project nor did he have access to any of our results, yet column 3 indicates a remarkably high correlation between his evaluation and the document relevance numbers. We then checked back to see how the documents containing information on "Thin Film" had been indexed (see col. 4). We found that the first five documents on our list had been indexed by both "Thin" and "Film". Three more documents had been indexed by "Film" alone, and other related terms. Two documents had not been indexed by either "Thin" or "Film", but only by a group of related terms, yet they contained information on "Thin Films" and had a high document relevance number. By using association factors, and a series of statistical steps, easily programmed for a computer, we were thus able to locate documents relevant to a request *even though the document had not been indexed by the terms used in the request*.

The basic 5 steps in our new retrieval method can be summarized as follows:

(1) Prepare a profile for each request term. This profile consists of terms that have been used with the request term and have an Association Factor greater than 1.

(2) Compare the profiles of each request term and select those terms which appear in all or in a given number of profiles. These terms are called *first generation terms*.

(3) Treat the first generation terms as request terms and repeat steps 1 and 2. The resultant terms which are not already request

## UNCLASSIFIED INFORMATION RETRIEVAL

terms or first generation terms are called *second generation terms*.

(4) Make a table of association factors for the expanded list of request terms. The sum of the Association Factors for each term is called its *weight*. This weight indicates the degree of association between that term and the complete request.

(5) Compare the list of expanded request terms with the index terms of each document in the collection and add the weights of the terms that match. The sum of the weights is called the *document relevance number*. This number is used to present the documents to the requester in the order of their probable relevance to the request.

Thus far, our experiments have been conducted on an existing collection of documents already indexed by a manual Uniterm Coordinate Index System. However, we believe that the really significant fact about our discoveries is their potential use in an all-machine document storage and retrieval system. Such a system could start with automatic encoding of natural language, as described by Luhn of IBM, [4] [5] and end with the presentation of abstracts of the desired documents. The results of "auto-encoding", which is a distinctive vocabulary representing the document, might be a more reliable basis for the statistical manipulation of our system than the whim of an indexer. For instance, when we searched for documents on the weatherproofing of fabrics, we missed one because it had been indexed by the terms "Comprehensive", "Study", "Weatherproofing" and no others. Only a search through the one hundred and seven documents on weatherproofing would have turned up this document. However, if the document had been "auto-encoded", the necessary number of distinctive terms would have been ensured.

The association factor would be useful in selecting incoming documents for dissemination to company engineers. It would form the bridge between the language of the engineers requirements and the language used in the document. Each requirement would be surrounded with a profile of terms based on those supplied by the engineer and supplemented by those automatically generated from the document collection. The index terms assigned to incoming documents would then be compared with these profiles to determine the degree to which they fulfilled the requirements.

Moreover, we should not only be able to provide an engineer with incoming material related to his requirements, of which there is bound to be too much, but also tell him which of those items contain some new information in his field (i. e., new to the document collection). This feat can be accomplished by comparing the profiles of the request terms derived from the established document collection with the profiles of the request terms derived from a group of in-



H. E. STILES

UNCLASSIFIED

coming documents. Terms appearing in high association with the request terms in the new profiles which were not associated in the established profiles are indicators of new and distinctive material. By treating these new words plus the original request terms as a new set of request terms, the documents containing the new and distinctive information can be found.

Further applications of the association factor are suggesting themselves daily. It is hoped that this presentation will stimulate further discussion and experimentation.

**DOCUMENTS RELATED TO "THIN FILM"**  
**ARRANGED BY DOCUMENT RELEVANCE NUMBERS**

1	2	3	4
Document Number	Document Relevance Numbers	Degree of Association	Use of Thin and/or Film as Index Terms
S-66,794	24.32	Yes	Thin Film
S-51,212	24.22	Yes	Thin Film
S-51,050	24.22	Yes	Thin Film
S-33,067	24.22	Yes	Thin Film
S-33,068	22.47	Yes	Thin Film
S-95,555	19.87	Yes	Film
S-34,019	15.59	Yes	Film
S-18,958	15.30	P	
S-73,671	14.83	Yes	
S-38,473	12.54	No	
S-37,438	11.81	M	
S-35,837	11.20	M	
S-39,631	10.72	M	
S-35,838	10.14	P	
S-65,855	10.08	M	
S-80,485	10.05	Yes	Film
S-76,529	9.83	Yes	
S-44,571	9.66	M	
S-56,755	9.50	M	
S-42,772	9.38	M	
S-63,862	9.38	No	
S-33,834	9.30	P	
S-33,835	9.30	P	
S-33,832	9.30	P	
S-35,839	9.30	P	
S-80,309	9.18	P	
S-59,129	9.12	P	

## UNCLASSIFIED INFORMATION RETRIEVAL

1	2	3	4
Document Number	Document Relevance Numbers	Degree of Association	Use of Thin and/or Film as Index Terms
S-70,145	8.66	M	
S-59,442	8.03	P	
S-60,834	7.95	M	
S-49,629	7.85	No	
S-71,275	7.27	No	
S-51,499	7.16	P	
S-33,831	7.13	M	
S-31,735	7.13	M	
S-66,513	6.93	P	
S-31,620	6.94	No	
S-31,620	6.94	No	
S-80,360	6.94	No	
S-61,700	6.91	No	
S-44,921	6.40	P	
S-59,130	6.40	P	
S-53,424	6.36	No	
S-78,885	6.03	M	
S-55,371	6.01	No	
S-38,474	5.82	P	
S-80,974	5.33	No	
S-48,093	5.23	No	
S-80,293	5.23	No	
S-55,644	4.65	No	
S-58,247	4.41	P	
S-60,114	4.41	P	
S-45,420	4.41	P	
S-28,975	4.26	No	
S-37,031	2.70	No	
S-71,296	2.70	No	

## REFERENCES

- [1] Sanford, Albert and Theriault, Frederic R. "Problems in the Application of Uniterm Coordinate Indexing" *College and Research Libraries*. Vol. 17 No. 1, January 1956. pp. 19-23.
- [2] Maron, M. E., Kuhns, J. L. and Ray, L. C. "Probabilistic Indexing" Ramo-Wooldridge, Data Systems Project Office, *Technical Memorandum No. 3*. June 1959.

H. E. STILES

UNCLASSIFIED

- [3] Yates, F., "Contingency Tables Involving Small Numbers and the Chi Square Test," *Supplement to the Journal of the Royal Statistical Society*, 1 (1934), pp. 217-235.
- [4] Luhn, H. P. *Potentialities of Auto-Encoding of Scientific Literature*, International Business Machines Corporation, Research Center, Research Report RC-101, May 15, 1959.
- [5] Luhn, H. P. "Auto-Encoding of Documents for Information Retrieval Systems" (In Boaz, Martha, ed. *Modern Trends in Documentation*. New York, Pergamon Press, 1959.)

~~CONFIDENTIAL~~

## TELEPHONE DIRECTORY OF CONTRIBUTORS TO THIS ISSUE

<i>Author</i>	<i>Office</i>	<i>Extension</i>
H. H. Campaigne	REMP	7249
W. F. Friedman*		
W. W. Jacobs	REMP	7249
P. E. Neff	CSEC	60501
H. E. Stiles	CREF	7267

---

\* Special Consultant

~~CONFIDENTIAL~~

CREF(AG)-Apr 61-MAT-65015D

245 FRIEDMANN P GS 13 R4 REMP NSA TECH JOURNAL

VOL 6 NO 1

## CERTIFICATE OF DESTRUCTION

Date of Destruction

(Signature of Destroying Officer)

(Signature of Witnessing Officer)

Mr William F Friedman

Introduction

to

Cryptology

Lectures I, II, III

~~CONFIDENTIAL~~

## An Introduction to Cryptology - I

BY WILLIAM F. FRIEDMAN

*Confidential*

*The first of series of lectures prepared by Mr. Friedman for delivery to an audience assumed to be totally unfamiliar with the subject.*

The objective of this series of lectures is to create an awareness of the background, development, and manner of employment of a science that is the basis of a vital military offensive and defensive weapon known as CRYPTOLOGY, a word that comes from the Greek *kryptos*, meaning *secret* or *hidden*, plus *logos*, meaning *knowledge* or *learning*. Cryptology will be specifically defined a little later; at the moment, however, I'm sure you know that it has to do with *secret communications*.

Let me say at the outset of these lectures that I may from time to time touch upon matters which are perhaps essentially peripheral or even irrelevant to the main issues, and if a defense is needed for such occasional browsing along the by-ways of the subject, it will be that long preoccupation with any field of knowledge begets a curiosity the satisfaction of which is what distinguishes the dedicated professional from the person who merely works just to gain a livelihood in whatever field he happens to find himself a job. That's not much fun, I'm afraid. By the way, a British writer, James Agate, defines a professional as the man who can do his job even when he doesn't feel like doing it; an amateur, as a man who can't do his job even when he does feel like doing it. This is pretty tough on the gifted amateur and I for one won't go all the way with Agate's definition. There are plenty of instances where gifted amateurs have done and discovered things to the chagrin and red-facedness of the professionals.

Coming back now to the main thoroughfare after the foregoing brief jaunt along a by-way, I may well begin by telling you that the science of cryptology has not always been regarded as a vital military offensive and defensive weapon, or even as a weapon in the first place. Here I am reminded of a story in a very old book on cryptography. The story is probably apocryphal, but it's a bit amusing, and I give it for what it's worth.

It seems that about two thousand years ago there lived a Persian queen named Semiramis, who took an active interest in cryptology. She was in some respects an extraordinarily unpleasant woman and we learn without surprise that she met with an untimely death. She left behind her instructions that her earthly remains were to be placed in a golden sarcophagus within an imposing mausoleum, on the outside of which, on its front stone wall, there was to be graven a message, saying:

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~ AN INTRODUCTION TO CRYPTOLOGY -I

Stay, weary traveller!  
 If thou art footsore, hungry, or in need of money--  
 Unlock the riddle of the cipher graven below,  
 And thou wilt be led to riches beyond all dreams of avarice!

Below this curious inscription was a cryptogram, a jumble of letters without meaning or even pronounceability. For several hundred years the possibility of sudden wealth served as a lure to many experts who tried very hard to decipher the cryptogram. They were all without success, until one day there appeared on the scene a long-haired, be-whiskered, and bespectacled savant who, after working at the project for a considerable length of time, solved the cipher, which gave him detailed instructions for finding a secret entry into the tomb. When he got inside, he found an instruction to open the sarcophagus, but he had to solve several more cryptograms the last one of which may have involved finding the correct combination to a 5-tumbler combination lock --who knows? Well, he solved that one too, after a lot of work, and this enabled him to open the sarcophagus, inside which he found a box. In the box was a message, this time in plain language, and this is what it said:

O, thou vile and insatiable monster! To disturb these poor bones!  
 If thou hadst learned something more useful than the art of  
 deciphering,  
 Thou wouldst not be footsore, hungry, or in need of money!

I'm frank to confess that many times during my 40-year preoccupation with cryptology, and generally near the middle and the end of each month, I felt that good old Queen Semiramis knew what she was talking about. However, earning money is only a part of the recompense for working in the cryptologic field, and I hope that most of you will find out sooner or later what some of these other recompenses are, and what they can mean to you.

If Queen Semiramis thought there are other things to learn that are more useful than the art of deciphering, I suppose we'd have to agree, but we are warranted in saying, at least, that there isn't any question about the importance of the role that cryptology plays in modern times: all of us are influenced and affected by it, as I hope to show you in a few minutes.

I shall begin by reading from a source which you'll all recognize--*Time*, the issue of 17 December 1945. I will preface the reading by reminding you that by that date World War II was all over -- or at least V-E and V-J days had been celebrated some months before. Some of you may be old enough to remember very clearly the loud clamor on the part of certain vociferous members of Congress, who had for years been in-

~~CONFIDENTIAL~~



sisting upon learning the reasons why we had been caught by surprise in such a disastrous defeat as the Japanese had inflicted upon us at Pearl Harbor. This clamor had to be met, for these Congressmen contended that the truth could no longer be hushed up or held back because of an alleged continuing need for military secrecy, as claimed by the Administration and by many Democratic senators and representatives. The war was over -- wasn't it? -- Republican senators and representatives insisted. There had been investigations—a half dozen of them—but all except one were *Top Secret*. The Republicans wanted—and at last they got what they desired—a grand finale Joint Congressional Investigation which would all be completely open to the public. No more secrets! It was spectacular. Not only did the Congressional Inquiry bring into the open every detail and exhibit uncovered by its own lengthy hearings, but it also disclosed to America and to the whole world everything that had been said and shown at all the previous Army and Navy investigations. Most of the information that was thus disclosed had been, and much of it still was *Top Secret*; yet all of these precious secrets became matters of public information as a result of the Congressional Investigation.

There came a day in the Congressional Hearings when the Chief of Staff of the United States Army at the time of the Pearl Harbor Attack, 5-star General George C. Marshall, was called to the witness stand. He testified for several long, long days, eight of them in all. Toward the end of the second day of his ordeal he was questioned about a letter it had been rumored he'd written to Governor Dewey in the Autumn of 1944, during the Presidential Campaign. The letter was about codes. With frozen face, General Marshall balked at disclosing the whole letter. He pleaded most earnestly with the Committee not to force him to disclose certain of its contents, but to no avail. He had to bow to the will of the majority of the Committee. I shall now read from *Time* a bit of information which may be new to many of my listeners, especially to those who were too young in December 1945 to be delving into periodical literature or to be reading any pages of the daily newspaper other than those on which the comics appear.

Said *Time*, and I quote:

"U S citizens discovered last week that perhaps their most potent secret weapon of World War II was not radar, not the VT fuse, not the atom bomb, but a harmless little machine which cryptographers had painstakingly constructed in a hidden room in Washington. With this machine, built after years of trial and error, of inference and deduction, cryptographers had duplicated the decoding devices used in Tokyo. Testimony before the Pearl Harbor Committee had already shown that the machine, known as '*Magic*' was in use long before December 7, 1941, and had given ample warning of the Japs' sneak attack, if only U S brass hats had been smart enough to realize it. Now, General Marshall continued the story of '*Magic*'s' magic

~~CONFIDENTIAL~~ AN INTRODUCTION TO CRYPTOLOGY -I

1. 'It had enabled a relatively small U.S. Force to intercept a Jap invasion fleet, win a decisive victory in the Battle of the Coral Sea, thus saving Australia and New Zealand.

3. 'It had directed U.S. submarines unerringly to the sea lanes where Japanese convoys would be passing.

2. 'It had given the U.S. full advance information on the size of the Jap forces advancing on Midway, enabled our Navy to concentrate ships which otherwise might have been 3,000 miles away, thus set up an ambush which proved to be the turning-point victory of the Pacific war

4. 'By decoding messages from Japan's Ambassador Oshima in Berlin, often reporting interviews with Hitler, it had given our forces invaluable information on German war plans'."

*Time* goes on to give more details of that story, to which I may later return but I can't leave this citation of what cryptology did toward our winning of World War II without telling you that the account given by *Time* of the achievements of *Magic* makes it appear that *all* the secret intelligence gained from our reading Japanese messages was obtained by using that "harmless little machine" which *Time* said was used in Tokyo by the Japanese Foreign Office. I must correct that error by explaining first that *Magic* was not the name of the machine but a term used to describe the intelligence material to which the machine, among other sources, contributes and then by telling you that the secret information we obtained that way had little to do with those portions of the *Magic* material which enabled our Navy to win such spectacular battles as those of the Coral Sea and Midway, and to waylay Japanese convoys. The naval parts of *Magic* were nearly all obtained from Japanese naval messages by our own very ingenious U.S. Navy cryptanalysts. At that time, I may tell those of you who are new, the Army and Navy had separate but cooperating cryptologic agencies and activities; the United States Air Force was not yet in existence as an autonomous and separate component of the Armed Forces, and work on Japanese, German, and Italian air-force communications was done by Army cryptanalysts, admirably assisted by personnel of what was then known as the Army Air Corps.

It is hardly necessary to tell you how carefully the *Magic* of World War II was guarded before, during, and after the war until the Congressional Inquiry brought most of it out in the open. Some remaining parts of it are still very carefully guarded. Even the fact of the existence of *Magic* was known to only a very few persons at the time of Pearl Harbor -- and that is an important element in any attempt to explain why we were caught by surprise by the Japanese at Pearl Harbor in a devastating attack that crippled our Navy for many months. Let me read a bit from page 261 of the Report of the Majority of the Joint Congressional Investigation of the attack:

~~CONFIDENTIAL~~

"The *Magic* intelligence was pre-eminently important and the necessity for keeping it confidential cannot be overestimated. However, so closely held and top secret was this intelligence that it appears that the *fact* that the Japanese codes had been broken was regarded as of more importance than the *information* obtained from decoded traffic."

*Time* says, in connection with this phase of the story of *Magic* during World War II

"So priceless a possession was *Magic* that the U S high command lived in constant fear that the Japs would discover the secret, change their code machinery, force U S cryptographers to start all over again "

Now I don't want to over-emphasize the importance of communications intelligence in World War II, but I think it warranted to read a bit more of what is said about its importance in the Report of the Majority. The following is from p. 232.

" all witnesses familiar with *Magic* material throughout the war have testified that it contributed enormously to the defeat of the enemy, greatly shortened the war, and saved many thousands of lives."

General Chamberlin, who was General MacArthur's operations officer, or G-3, throughout the war in the Pacific, has written: "The information G-2, that is, the intelligence staff, gave me in the Pacific Theater alone saved us many thousands of lives and shortened the war by no less than two years." We can't put a dollars-and-cents value on what our possession of COMINT meant in the way of saving lives, but we can make a dollars-and-cents estimate of what communications intelligence meant by shortening the war by two years, and the result of that estimate is that it appears that \$1.00 spent for that sort of intelligence was worth \$1,000 spent for other military activities and materials

In short, when our commanders had that kind of intelligence in World War II they were able to put what small forces they had at the right place, at the right time. But when they didn't have it--and this happened, too, --their forces often took a beating. Later on we'll note instances of each type.

I hope I've not tried your patience by such a lengthy preface to the real substance of this series of lectures, let's get down to brass tacks. For those of you who come to the subject of cryptology for the first time, a few definitions will be useful, in order that what I shall be talking about may be understood without question. Agreement on basic terminology is always desirable in tackling any new subject. In giving you the definitions there may be a bit of repetition because we shall be looking at the same terms from somewhat different angles.

~~CONFIDENTIAL~~ AN INTRODUCTION TO CRYPTOLOGY - I

First, then, what is cryptology? Briefly, we may define it as the doctrine, theory, or branch of knowledge which treats of hidden, disguised, or secret communications. You won't find the word in a small dictionary. Even Webster's Unabridged defines it merely as "secret or enigmatical language", and in its "Addenda Section", which presumably contains new or recently-coined words, it is defined merely as "the study of cryptography". Neither of these definitions is broad or specific enough for those who are going to delve somewhat deeply into this science.

Cryptology has two main branches; the first is cryptography, or, very briefly, the science of preparing secret communications, and the second is cryptanalysis, or the science of solving secret communications. Let's take up cryptography first, because as a procedure it logically precedes cryptanalysis: before solving anything there must be something to solve.

Cryptography is that branch of cryptology which deals with the various means, methods, devices, and machines for converting messages in ordinary, or what we call plain language, into secret language, or what we call cryptograms. Here's a picture of one of the most famous cryptograms in history. It was the solution of this cryptogram which resulted in bringing America into World War I on the side of the Allies on 6 April

**WESTERN UNION**  
**TELEGRAM**

via Galvesto

JAN 18 1917

GERMAN LEGATION  
MEXICO CITY

130	13042	134C	8501	115	3528	416	17214	6481	11310
1814	18222	2140C	13247	11518	25877	13005	3494	14936	
8092	59C5	11311	11392	10371	0302	21290	5161	39095	
300	1504	11200	18200	18101	0517	0228	17694	4473	
2224	22200	10402	21589	07893	5409	13918	8986	12137	
1233	4725	4458	59.5	17108	13851	4458	17149	14471	0706
1385	2224	0929	14991	7382	1585	07893	14218	36477	
000	1000	0000	5000	5454	16 C2	15217	22801	17158	
1001	0302	7110	2008	18222	0719	14531	15021	23846	
01	0000	22000	21604	4007	0407	2240	20853	4377	
2300	18140	22200	5900	13349	20420	39688	13732	80487	
000	0000	000	02202	1340	22049	13339	11286	22296	
10439	14014	4178	0992	8784	7632	7357	6926	52262	11.07
2110	21272	9340	9550	20004	15874	18502	18500	18807	
2180	5300	0381	98092	16127	13486	9350	9220	70088	14218
5144	2831	17920	11047	17148	11284	7687	7762	18388	9138
10482	97550	3569	3870						

DEPLSTOP??

Charge German Legation

Fig 1 - The Zimmerman Telegram

~~CONFIDENTIAL~~

1917, just about six weeks after it was solved. I'll tell you about it later in this series.

Cryptography also includes the business of reconverting the cryptograms into their original plain-language form, by a direct reversal of the steps followed in the original transformation. This implies that the persons involved in both of these bits of business, those at the enciphering and sending end, and those at the receiving and deciphering end, have an understanding as to what procedures, devices, and so on, will be used and exactly how--down to the very last detail. The what and the how of the business constitutes what is generally referred to as the *key*. The key may consist of a set of rules, alphabets, procedures, and so on; it may also consist of an ordinary book which is used as a source of keys; or it may be a specialized book, called a *code book*. That cryptogram I just showed you was made by using a book--a German codebook.

To *encrypt*, is to convert or transform a plain-text message into a cryptogram by following certain rules, steps, or processes constituting the key or keys and agreed upon in advance by the correspondents, or furnished them by higher authority.

To *decrypt* is to reconvert or to transform a cryptogram into the original equivalent plain-text message by a direct reversal of the encrypting process that is, by applying to the cryptogram the key or keys, usually in a reverse order, employed in producing it.

A person who encrypts and decrypts messages by having in his possession the necessary keys, is called a *cryptographer*, or a *cryptographic clerk*.

Encrypting and decrypting are accomplished by means collectively designated as *codes and ciphers*. Such means are used for either or both of two purposes (1) secrecy, and (2) economy. Secrecy usually is far more important in diplomatic and military cryptography than economy, but it is possible to combine secrecy and economy in a single system. Persons technically unacquainted with cryptology often talk about "cipher codes", a term which I suppose came into use to differentiate the term "code" as used in cryptology from the same term as used in other connotations, as, for example, the Napoleonic Code, a traffic code, a building code, a code of ethics, and so on. Now, in cryptology, there is no such thing as a "cipher code". There are *codes* and there are *ciphers*, and we might as well learn right off the differences between them, so that we get them straightened out in our minds before proceeding further.

In ciphers, or in cipher systems, cryptograms are produced by applying the cryptographic treatment to individual letters of the plain-text messages, whereas, in codes, or in code systems, cryptograms are produced by applying the cryptographic treatment generally to entire words, phrases, and sentences of the plain-text messages. More specialized meanings of the terms will be explained in detail later, but in a moment I'll show you an example of a cryptogram in cipher and one in code.

A cryptogram produced by means of a cipher system is said to be in *cipher* and is called a *cipher message*, or sometimes, simply a *cipher*. The act or operation of encrypting a cipher message is called *enciphering*, and the enciphered version of the plain text, as well as the act or process itself, is often referred to as the *encipherment*. A cryptographic clerk who performs the process serves as an *encipherer*. The corresponding terms applicable to *decrypting* cipher messages are *deciphering*, *decipherment*, *decipherer*.

A cryptogram produced by means of a code system is said to be in *code*, and is called a *code message*. The text of the cryptogram is referred to as *code text*. This act or operation of encrypting is called *encoding*, and the encoded version of the plain text, as well as the act or process itself, is referred to as the *encodement*. The clerk who performs the process serves as an *encoder*. The corresponding terms applicable to the decrypting of code messages are *decoding*, *decodement*, and *decoder*. A clerk who encodes and decodes messages by having in his possession the pertinent code books is called a *code clerk*.

Technically, there are only two distinctly different types of treatment which may be applied to written plain text to convert it into a cipher, yielding two different classes of ciphers. In the first, called *transposition*, the letters of the plain text retain their original identities and merely undergo some change in the relative positions, with the result that the original text becomes unintelligible. Here's an authentic example of a transposition cipher, I call it authentic because it was sent to President

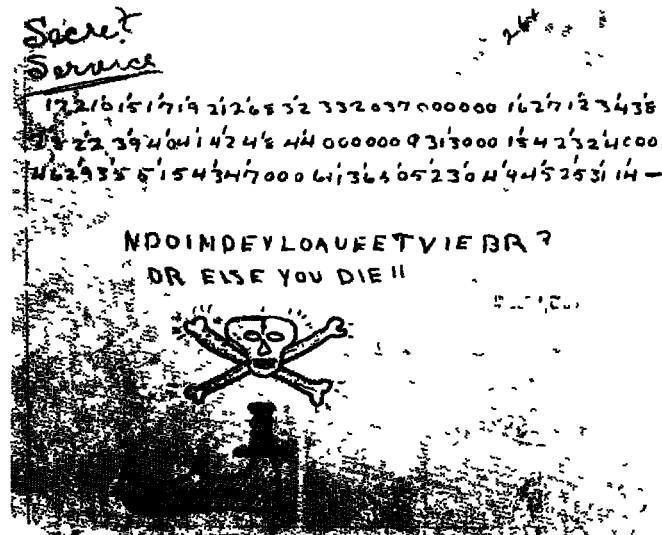


Fig 2

Roosevelt and the Secret Service asked me to decipher it. Imagine my chagrin when I had to report that it says "Did you ever bite a lemon?" In the second, called *substitution*, the letters of the plain text retain their original relative positions, but are replaced by other letters with different sound values, or by symbols of some sort, so that the original text becomes unintelligible.

Nobody will quarrel with you very hard if you wish to say that a code system is nothing but a specialized form of substitution; but it's best to use the word "code" when a code book is involved, and to use "substitution cipher" when a literal system of substitution is used.

It is possible to encrypt a message by a substitution method and then to apply a transposition method to the substitution text, or vice versa. Combined transposition-substitution ciphers do not form a third class of ciphers; they are only occasionally encountered in military cryptography. Applying a cipher to code groups is a very frequently used procedure and we'll see cases of that too.

Here's an example of a substitution cipher, and a very simple one. It was found on a German spy in World War II. Here's the cipher alphabet; here's the plain text which happened to be in German; and here's

### RUMRICH SPY CASE

(GUNTHER GUSTAVE RUMRICH ET AL.)

PHOTOGRAPHIC COPY OF THE CIPHER  
WITH THE GERMAN AND ENGLISH  
DECIPHERING INTERLINED

4-417--455-4 2679 0-79 611  
LIEBE JENNIE NACH MISS MOOS  
DEAR JENNIE PLEASE GO TO MISS MOOS  
1-79 1-004 16 112-127-11-1-9-7442  
GEM BITTE AN ABFAHRTSTAG ZWISCHEN  
ON THE DAY OF DEPARTURE BETWEEN  
3-6211 1155-40 741 1-5211 8-740  
3-5 UHR DANN IST DER DOKTOR NICHT  
3-5 O'CLOCK THEN THE DOCTOR WILL NOT BE  
11 741 4-11 1155 741 741 8-7904  
DA DER SOLL VON DER SACHE NIGHTS  
THERE HE IS NOT TO KNOW ANYTHING ABOUT THE  
9-7945 1-4-8 5211

WISSEN DEIN KARL  
MATTER YOUR KARL

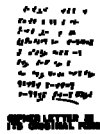


Fig. 3

the cipher text or encipherment.

Now for an example of a cryptogram in code. On the following page is a plain-text message in the handwriting of President Wilson, to his special emissary in London, Colonel House. Also contained on the next page is the cryptogram after the plain text was encoded by Mrs. Wilson. The President then himself typed out the final message on his own typewriter, for transmission by the Department of State. It would appear that President Wilson lacked confidence in the security of the Department of State's methods--and maybe with good reason, as may be seen in the following extract from a letter dated 14 September 1914 from the President to Ambassador Page in London: "We have for some time been trying to trace the leaks, for they have occurred frequently, and we are now convinced that our code is in possession of persons at

~~CONFIDENTIAL~~ AN INTRODUCTION TO CRYPTOLOGY - I

*Memorandum (Honor)*

*It now looks as if our  
several difficulties with  
Germany would be presently  
adjusted. So soon as they  
are the German line separ-  
ately from the Soviet will be  
information that in force Eng-  
land to make at least a  
great concession to our un-  
assailable claims of right.  
This is just what I  
saw this for your informa-  
tion and guidance*

*H.*

Fig 4

39608-33391-37200-

67906-32040-2284-52927

12726-2041K-65092-29004-72610

30885-68613-54088-43336-49674

46352-22643-65062-42217-17802

47156-220210-36818-66908-49733

58436-17288-16137-57957-32786

24556-17503-39195-44120-42630

22662-17686-47124-41126-70104

44885-

Fig 5

intermediary points. We are going to take thoroughgoing measures." Perhaps one of the measures was that the President got himself a code of his own. I must follow this up some day.

A cipher device is a relatively simple mechanical contrivance for encipherment and decipherment, usually "hand-operated", or manipulated by the fingers, as for example, a device with concentric rings of alphabets, manually powered. On the next page is an example -- a cipher device with such rings. I'll tell you about it later. A cipher machine is a relatively complex apparatus or mechanism for encipherment and decipherment, usually equipped with a typewriter keyboard and generally requiring an external power source. Modern cryptology, following the trend in mechanization and automation in other fields, now deals largely with cipher machines, some highly complicated. Also pictured on next page is a modern cipher machine with keyboard and printing mechanism.

One of the expressions which unformed laymen use, but which you must never use, is "the German code", or "the Japanese code", or "the Navy cipher", and the like. When you hear this sort of expression you may put the speaker down at once as a novice. There are literally hundreds of different codes and ciphers in simultaneous use by every large and important government or service, each suited to a special purpose, or where there is a multiplicity of systems of the same general nature, the object is to prevent a great deal of traffic being encrypted in the same key, thus overloading the system and making it vulnerable to

~~CONFIDENTIAL~~



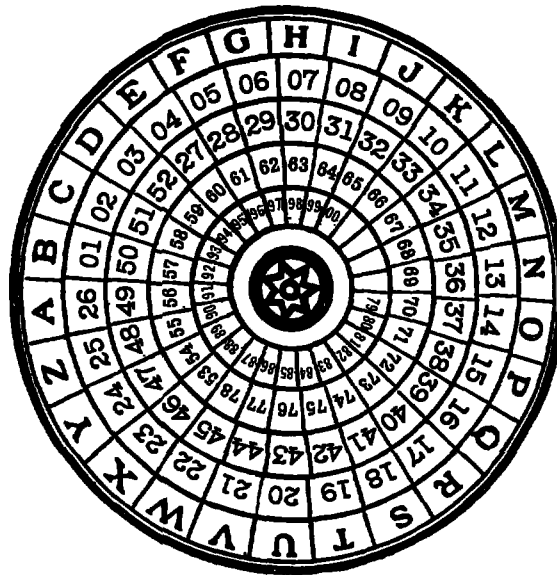


Fig 6

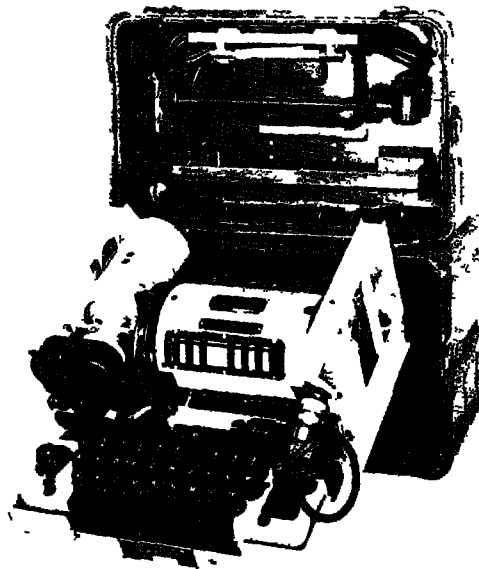


Fig. 7 - TSEC/KL-7 Cipher Machine (U S.)

~~CONFIDENTIAL~~ AN INTRODUCTION TO CRYPTOLOGY - I

attack by methods and procedures to be mentioned in broad terms in a few moments.

The need for secrecy in the conduct of important affairs has been recognized from time immemorial. In the case of diplomacy and organized warfare this need is especially important in regard to communications. However, when such communications are transmitted by electrical means, they can be heard or, as we say, *intercepted*, and copied by unauthorized persons, usually referred to collectively as *the enemy*. The protection resulting from all measures designed to deny to the enemy information of value which may be derived from the interception and study of such communications is called *communication security*, or, for short, COMSEC.

In theory, any cryptosystem except one, to be discussed in due time, can be attacked and "broken", i.e., solved, if enough time, labor, and skill are devoted to it, and if the volume of traffic in that system is large enough. This can be done even if the general system and the specific key are unknown at the start. You will remember that I prefaced my statement that any cryptosystem can be solved by saying "*in theory*," because in military operations theoretical rules usually give way to practical considerations.

That branch of cryptology which deals with the principles, methods, and means employed in the *solution* or *analysis* of cryptosystems is called *cryptanalytics*. The steps and operations performed in applying the principles of cryptanalytics constitute *cryptanalysis*. To *cryptanalyze* a cryptogram is to solve it by cryptanalysis. A person skilled in the art of cryptanalysis is called a *cryptanalyst*, and a clerk who assists in such work is called a *cryptanalytic clerk*.

Information derived from the organized interception, study, and analysis of the enemy's communications is called *communication intelligence*, or, for short, COMINT. Let us take careful note that COMINT and COMSEC deal with communications. Although no phenomenon is more familiar to us than that of communication, the fact of the matter is that this magic word means many things to many people. A definition of communication that is broad enough for our purposes would be that communication deals with intelligent *messages* exchanged between intelligent beings. This implies that human beings, and human operators are involved in the preparation, encryption, transmission, reception, decryption, and recording of messages which at some stage or stages are in written form and in some stage or stages are in electrical form as signals of one sort or another. But in recent years there have come into prominence and importance electrical signals which are not of the sort I've just indicated. They do not carry "messages" in the usual sense of the word, they do not convey from one human being to another an intelligible sequence of words and an intelligible sense. I refer here to electrical or electronic signals such as are employed in homing or directional beacons, in radar, in tele-

~~CONFIDENTIAL~~

metering or recording data of an electrical or electronic nature at a distance, and so on. Information obtained from a study of enemy electronic emissions of these sorts is called *electronic intelligence*, or, for short, ELINT. The particular or specialized study of enemy radar signals is called RADINT. All these, COMINT, ELINT, RADINT comprise SIGINT, that is, *signal intelligence*. Cryptology is the science which is concerned with *all* these branches of secret signalling.

In this series of lectures we shall be concerned only with COMSEC and COMINT, leaving for others and for other times the subjects of ELINT, RADINT, and so on. This means that we shall deal with communications or *messages*

Communication may be conducted by any means susceptible of ultimate interpretation by one of the five senses, but those most commonly used are seeing and hearing. Aside from the use of simple visual and auditory signals for communication over relatively short distances, the usual method of communication between or among individuals separated from another by relatively long distances involves, at one stage or another, the act of writing or of speaking over a telephone.

Privacy or secrecy in communication by telephone can be obtained by using equipment which affects the electrical currents involved in telephony, so that the conversations can be understood only by persons provided with suitable equipment properly arranged for the purpose. The same thing is true in the case of facsimile transmission (i.e., the electrical transmission of ordinary writing, pictures, drawings, maps). Even today there are already simple forms of enciphered television transmissions. Enciphered facsimile is called CIFAX, enciphered telephony, CIPHONY, and enciphered television, CIVISION. However, these lectures will not deal with these electrically and cryptanalytically more complex forms of cryptology. We shall stick to enciphered or encrypted writing--which will be hard enough for most of us.

Writing may be either visible or invisible. In the former, the characters are inscribed with ordinary writing materials and can be seen with the naked eye; in the latter, the characters are inscribed by means or methods which make the writing invisible to the naked eye. Invisible writing can be prepared with certain chemicals called sympathetic or secret inks, and in order to "develop" such writing, that is, make it visible, special processes must usually be applied. Shown on the next page is an interesting example--the developed secret-ink message that figured in an \$80,000,000 suit won by two American firms against the German Government after World War I sabotage was proved. There are also methods of producing writing which is invisible to the naked eye because the characters are of microscopic size, thus requiring special microscopic and photographic apparatus to enlarge such writing enough to

CONFIDENTIAL AN INTRODUCTION TO CRYPTOLOGY - I

## The Yukon Trail

## THE BLUE

and thinking you will be a good man to  
 "His narrowed eyes held a cold glit-  
 "You must know he is innocent. You  
 must—"

"I know only what the evidence  
 shows," he cut in, warily on his guard  
 "He may or may not have been one of  
 my attackers from the first blow I  
 was dazed. But everything points to  
 that he hired."  
 "Oh, no," interrupted the Irish girl,  
 her dark eyes shining softly. "The  
 way of it is that he saved your life, that  
 he fought for you and that he is in  
 prison because of it."

"If that is true, why doesn't he bring  
 some proof of it?"

"Proof," she cried scornfully. "Be-  
 tween friends—"

"He's no friend of mine. The man  
 is a meddler. I despise him."

"And I am liking him very much,"  
 she flung back stanchly.

Macdonald looked up at the vivid,  
 flushed face and found it wholly charm-  
 ing. He liked her none the less because  
 her fine eyes were hot and defiant in  
 defiance of his will.

"Oh, well," he smiled. "I'll let him  
 out if you'll do me a good turn too."

"Thank you. It's a bargain."

"Then sing to me."

She moved to the piano and that  
 "I'll sing."

"Sing 'Divided'."

Her long lashes veiled her soft eyes  
 while she considered. In a way he had  
 picked her into singing for him a love-  
 song she did not want to sing. But she  
 made no protest. Swiftly she turned  
 and sat along the bench. Her fingers  
 touched the keys and she began.

He watched the beauty and warmth  
 of her dainty youth with eyes that mir-  
 rored the hunger of his heart. How  
 gloriously she carried her dusky hair  
 back! With what a gallant spirit she did  
 all things! He was usually a crafty man,  
 but when he was with her, it seemed to  
 him as if God had thrown the person-  
 ality of all sorts of brave, fine promises.

She paid her pledge in full. After  
 the first two stanzas were finished  
 she sang the last ones more slowly.

An whirl about the weather when  
 I'd have could I do it."

Is it me that would be wined to  
 "Grip the oars and go along it."

Oh, I could tug and by the light  
 Of day or moon or star

But there's colder things than salt waves  
 Between us, so they are  
 Ours alone!

Sure, well I know he'll never have  
 The heart to come to me,  
 An' love is wild as any ave

That wanders on the sea  
 'Tis the same if I'm near me,  
 'Tis the same if he is far

His thoughts are hard an' ever hard  
 Between us, so they are  
 Ours alone!

Her hands dropped from the keys,  
 and she turned slowly on the end of  
 the seat. The dark lashes fell to her  
 not cheeks. He did not speak, but she  
 felt the steady insistence of his gaze.

In self-defense she looked at him.  
 The color of his face lent accent to  
 the fire that smoldered in his eyes.

"The going to marry you, Sheba  
 Make up your mind to that girl," he  
 said harshly.

There was infinite pity in the look  
 she gave him. "There's colder things  
 than salt waves between us, so they  
 are," she quoted.

"Not if I love you and you love me  
 I'll trample down everything that comes  
 between us."

He swung to a sitting position on the  
 lounge. Through the steel-gray eyes in  
 the hooding face his masterful spirit  
 wrestled with hers.

lean-shouldered, with head, powerful should-  
 ers and deep chest, he dominated his  
 world ruthlessly. But the slim, Irish  
 girl held her own.

"Must we go through that again?"  
 she asked gently.

"Again and again until you see  
 reason."

She knew the tremendous driving  
 power of the man, and she was afraid  
 in her heart that he would sweep her  
 from the moorings to which she clung.

There is something like I have  
 told you."

The embarrassed blush  
 lifted bravely from the flushed cheeks  
 to meet steadily his look. "I don't think  
 that I—care for you."

"Is that am  
 richness. But  
 I don't—not with the full of my heart."

Fig. 8

CONFIDENTIAL

W. F. FRIEDMAN

-CONFIDENTIAL-

make it visible to the naked eye. Here's an example--a code message in a space not much larger than the head of a pin. A simple definition of secret writing would be to say that it comprises invisible writing and unintelligible visible writing.

There is one additional piece of basic information which it is wise to call to your attention before we proceed much further, and I'll begin by stating that the greatest and the most powerful instrument or weapon ever forged and improved by man in his long struggle for emancipation from utter dependence upon his own environment is the weapon of literacy--a mastery of reading and writing; and the most important invention, the one that made the weapon of literacy *practical*, was the invention of the *alphabet*. It is therefore a rather striking anomaly that we should now come to the study of another weapon--a counter-weapon to the weapon of literacy--the weapon of *secrecy*, the basic intent of which is to thwart the weapon that man struggled so long to forge. Secrecy is applied to make writing more difficult and the reading of of the writing very difficult, if not impossible.

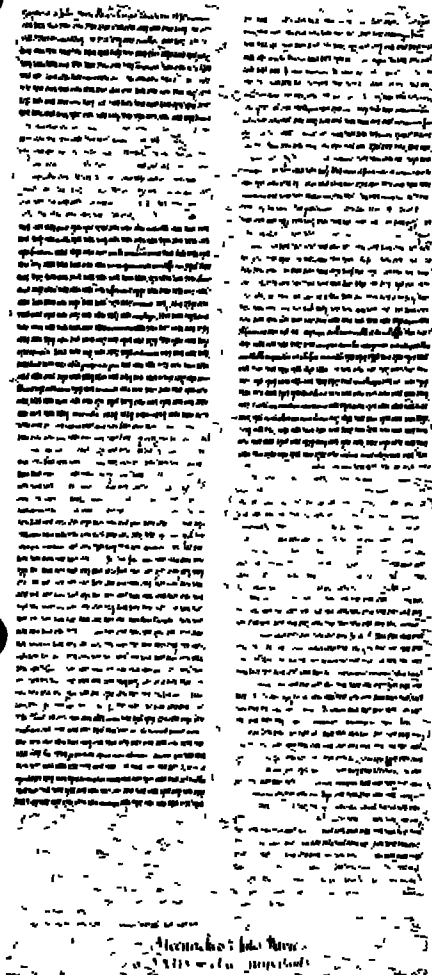


Fig. 9

Perhaps this is a good place to do a bit of theorizing about this matter of secrecy and what it implies.

Every person who enciphers a piece of writing, a message, or a text of any kind, for the purpose of hiding something or of keeping something secret, does so with the idea that some other person, removed from him in distance, or time, or both, is intended to decipher the writing or message and thus uncover the secret which was so hidden. A person may possess a certain piece of knowledge which he does not wish to forget, but which he is nevertheless unwilling to commit to open writing, and therefore he may jot it down in cryptic form for himself to decipher later, when or if the information is needed. The most widely known

~~CONFIDENTIAL~~ AN INTRODUCTION TO CRYPTOLOGY - I

example of such a cryptogram is found in Edgar Allan Poe's romantic tale *The Gold Bug*. That sort of usage of cryptography, however, is unusual. There are also examples of the use of cipher writing to establish priority of discovery, as did the astronomers Galileo and Huygens. Here's a picture which shows both examples. I suppose I should at least mention another sort of cryptic writing famous in literary history, the diaries of persons such as Samuel Pepys and William Byrd. These are commonly regarded as being "in cipher", but they were actually written in a more or less private shorthand and can easily be read without the help of cryptanalysis. On the next page is a page of Pepys diary.

Now there can be no logical reason, point, or purpose in taking the time and trouble to encipher anything unless it is expected that some other person is to decipher the cipher some time in the future. This means that there must exist some very direct, clear-cut and unambiguous relationship between the enciphering and deciphering operations. Just what such a relationship involves will be dealt with later, but at this moment all that it is necessary to say is that in enciphering there must be rules that govern or control the operations, that these rules must admit of no uncertainty or ambiguity, and that they must be susceptible of being applied with undeviating precision, since otherwise it will be difficult or perhaps impossible for the decipherer to obtain the correct answer when he reverses the processes or steps followed in the encipherment. This may be a good place to point out that a valid or authentic cryptanalytic solution cannot be considered as being merely what the cryptanalyst thinks or says he thinks the cryptogram means, nor does the solution represent an *opinion* of the cryptanalyst. Solutions are valid only insofar as they are objective and susceptible of demonstration or proof employing scientifically acceptable methods or procedures. It should hardly be necessary to indicate that the validity of the results achieved by cryptanalytic studies of authentic cryptograms rests upon the same sure and well-established scientific foundations, and is reached by the same sort of logic as are the discoveries, results, or "answers"

Translation of an extract from Galileo's letter to Kepler,  
published in Kepler's *DIOPHICUS* (1611)

"It is time for me to disclose the method of reading the letters which some weeks since I sent you as an enigma. It is time now, I mean, after I have become quite certain about the matter, so much so that I have no longer even a shadow of doubt.  
The words which I sent with their letters transposed, and which said,  
Hoc lunare a me Jove frustra leguntur, o J  
When reduced to their proper order, read thus,  
Cynthiae figurae assumuntur inter aurum  
The mother of the Loves rivals the phases of Cynthia that is,  
Jove's imitates the phases of the Moon"

Translation of an extract from Huygens' *SYSTEMA SATURNIUM*  
(1659)

"And this is that very hypothesis which, in the year 1656, on the 23th day of March, I put forth in confused letters together with my observation on the Saturnian Moon.  
Now the letters were a a a a a a a c c c c d  
e e e e h i i i i i i i l l l l l l u u u u u u u  
n n n p p p r r r t t t t t u u u u, which, being restored to their proper places, signify the following  
Anulus cingit, tellus, planus, aequum  
conspicere, ad eclipticam inclinato  
(It is encircled by a ring, thin plane,  
somewhat flattened, inclined to the ecliptic)

Fig. 10



~~CONFIDENTIAL~~—AN INTRODUCTION TO CRYPTOLOGY - I

achieved by any other scientific studies, namely observation, hypothesis, deduction, induction, and confirmatory experiment. Implied in what I have just said is the tacitly understood and now rarely explicitly stated assumption that two or more, equally competent and, if necessary, specially qualified investigators, each working independently upon the same material, will achieve identical or practically identical results.

Cryptology is usually and properly considered to be a branch of mathematics, although Francis Bacon considered it also a branch of grammar and what we now call linguistics. Mathematical and statistical considerations play an ever-increasing and prominent role in practical cryptology, but don't let my statement of this point frighten those of you who have not had much formal instruction in these subject. We have excellent cryptologists who have never studied more than arithmetic, and some of our best ones would hide if you were to go searching for mathematicians around here. What is needed is the ability to reason logically, as the mathematician sometimes does, and this ability is found in the most curious sorts of persons and places. So those of you who are frightened by the words mathematics and statistics take heart--you're not nearly so badly off as you may fear.

But now to return to the main theme, the place mathematics occupies in cryptology, let me say that just as the solution of mathematical problems leaves no room for the exercise of divination or other mysterious mental or psychic powers, so a valid solution to a cryptogram must leave no room for the exercise of such powers. In cryptologic science there is one and only one valid solution to a cryptogram, just as there is but one correct solution or "solution set" to any problem in mathematics. But perhaps I've already dwelt on this point too long, in any case, we'll come back to it later, when we come to look at certain types of what we may call pseudo-ciphers.

In the next lecture I'm going to give you a brief glimpse into the background or history of cryptology, which makes a long and interesting story that has never been told accurately and in detail. The history of communications security, that is, of cryptography, and the history of communications intelligence, that is, of cryptanalysis, which are but opposite faces of the same coin, deserve detailed treatment, but I am dubious that this sort of history will ever be written because of the curtain of secrecy and silence which officially surrounds the whole field of cryptology. *Authentic* information on the background and development of these vital matters having to do with the security of a nation is understandably quite sparse.

But in the succeeding lectures I'll try my best to give you authentic information, and where there's conjecture or doubt I'll so indicate. I must add, however, that in this series I'm going to have to omit many highly-interesting episodes and bits of information, not only because these

~~CONFIDENTIAL~~



lectures are of low classification, but also because we won't and can't for security considerations, go beyond a certain period in cryptologic history. Nevertheless, I hope you won't be disappointed, and that you'll learn certain things of great interest and importance, things to remember if you wish to make cryptology your vocation in life.

~~CONFIDENTIAL~~

## An Introduction to Cryptology—II

BY WILLIAM F FRIEDMAN

*Confidential*

*In this lecture, the author describes the earliest attempts at cryptography—from the invention of the art of writing to Bacon's "Bi-literarie" cipher.*

As I said at the close of the preceding lecture, a bit of history is always useful in introducing a subject belonging to a special and not too well known field; therefore, I'll proceed with some historical information about cryptology, which, as you learned before, comprises two closely related sciences, namely, cryptography and cryptanalysis. I will repeat and emphasize that they are but opposite faces of the same valuable coin; progress in one inevitably leads to progress in the other, and to be efficient in cryptology you must know something about each of them.












































Cryptography and cryptanalysis probably go back to the dawn of the invention and development of the art of writing itself. In fact, there is reason for speculating as to which came first—the invention of writing or the invention of cryptography; it's somewhat like the question as to which came first—the hen or the egg. It is possible that some phases of cryptography came before the art of writing had advanced very far.
















I've mentioned the art of writing. As in the case of other seemingly simple questions, such as, "why is grass green?", when we are asked to define writing we can't find a very simple answer, just because the answer isn't at all simple. Yet, Breasted, the famous University of Chicago historian and Orientalist, once said: "The invention of writing and of a convenient system of records on paper has had a greater influence in uplifting the human race than any other intellectual achievement in the career of man." There has been, in my humble opinion, no greater invention in all history. The invention of writing formed the real beginning of civilization. As language distinguishes man from other animals, so writing distinguishes civilized man from barbarian. To put the matter briefly, writing exists only in a civilization and a civilization cannot exist without writing. Let me remind you that animals and insects do communicate—there's no question about that; but writing is a thing peculiar to and found only as a phenomenon in which man and no animal or insect engages, and let's never forget this fact. Mankind lived and functioned for an

~~CONFIDENTIAL~~

enormous number of centuries before writing was discovered and there is no doubt that writing was preceded by articulate speech for eons—but civilization began only when men got the idea of and invented the art of writing. So far as concerns Western or Occidental civilization, writing is in essence, a means of representing the sounds of what we call speech or spoken language. Other systems of writing were and some still are handicapped by trying to represent things and ideas by pictures. I'm being a bit solemn about this great invention because I want to impress upon you what our studies in cryptology are really

## THE MAGIC

Good : as u  er  u'll  --w now  
u  know--t  every  pre  2 U a  or  
g8may 2 go th : a  of oppor  4 which u  as  
must f~~ond~~ the .  the s d/. w  o/d, l  a  
spacious  of t  have ill 2 c it, here the  s  
a  in  ing  ld.  th<sup>808</sup> who nter here  
may c  hem  as t  come 2 B. They c the k~~ey~~  
of f~~ree~~ they could  come, they    
kind of  whose  x  th  gentle,   
f~~ri~~ness; who know the  of the  of   
f~~ri~~ : who  r~~r~~ new /pe, new  age & nu  
ch  2 whom they  come n~~g~~.

This  (as soon, no d , u'll c) is o  b  
a magic  -- a litl .  hard 2 find,  1 2 w   
 371  R  & never  never reach: the  en  of  
gracious  s~~o~~.

The  we s  with  ey 2 nearby fri   
the  ds t  we b  or wave @   2  fr  
as  ab  -- th   us 2 the st/ of  t  
lii  the  t~~o~~ ly  u & me.  we f~~ond~~ the 

Fig. 1.

intended to do, namely, to defeat the basic or intended purpose of that great invention: instead of recording things and ideas for the dissemination of knowledge, we want and strive our utmost to prevent this aim from being realized, *except among our own brethren and under certain special circumstances*, for the purpose of our mutual security, our self-preservation And that's important.

Writing is a comparatively new thing in the history of mankind. No complete system of writing was used before about 3500 B.C.

Ordinary writing, the sort of writing you and I use, is perhaps an outgrowth or development of picture writing or rebus writing, which I'm sure most of you enjoyed as children. A rebus contains features of both ordinary and cryptographic writing; you have to "decrypt" the significance of some of the symbols, combine single letters with syllables, pronounce the word that is represented by pictures, and so on. Figure 1 is an example which I have through the courtesy of the Bell Telephone Laboratories. See how much of it you can make out in half a minute.

From rebus writing there came in due course alphabetic writing and let me say right now that the invention of the alphabet, which apparently happened only once in the history of mankind, in some Middle East Semitic region, in or near the Palestine-Syria area, then spread throughout the whole of the European continent, and finally throughout most of the world, is Western man's greatest, most important, and most far-reaching invention because it forms the foundation of practically all our written and printed knowledge, except that in Chinese. The great achievement of the invention of the alphabet was certainly not the creation of the signs or symbols. It involved two brilliant ideas. The first was the idea of representing merely the *sounds* of speech by symbols, that is, the idea of what we may call *phoneticization*, the second was the idea of adopting a system in which, roughly speaking, each speech sound is denoted or represented by one and only one symbol. Simple as these two ideas seem to us *now*, the invention was apparently made, as I've said, only once and the inventor or inventors of the alphabet deserve to be ranked among the greatest benefactors of mankind. It made possible the recording of the memory of mankind in our libraries, and from that single invention have come all past and present alphabets. Some of the greatest of men's achievements we are now apt to take for granted; we seldom give them any thought. The invention of the art of writing and the invention of the alphabet are two such achievements and they are worth pondering upon. Where would we be without them? Note that among living languages Chinese presents special problems not only for the cryptologist but also for the Chinese themselves. No

~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

Sinologist knows all the 80,000 or so Chinese symbols, and it is also far from easy to master merely the 9,000 or so symbols actually employed by Chinese scholars. How far more simple it is to use only 20 to 26 symbols! Being a monosyllabic language, it seems almost hopeless to try to write Chinese by the sort of mechanism used in an alphabetic polysyllabic language, attempts along these lines have been unsuccessful and the difficulties in memorizing a great many Chinese characters account for the fact that even now only about 10% of the Chinese people can read or write to any significant degree. The spread of knowledge in China is thereby much hampered.

We find instances of ciphers in the Bible. In Jeremiah Chapter 25, Verse 26 occurs this expression "And the King of Sheshakh shall drink after them." Also, again in Jeremiah 51:41 "How is Sheshakh taken!" Well, for perhaps many years that name "Sheshakh" re-

Jeremiah 25 : 26

"... and the king of Sheshakh shall drink after them."

Jeremiah 51 : 41

"How is Sheshakh taken! ... how is Babylon become an astonishment among the nations!"

11	10	9	8	7	6	5	4	3	2	1
Kh	I	T	Ch	Z	V	H	D	G	B	A
כ	י	ט	ח	ז	ו	ה	ד	ג	ב	א
ל	מ	נ	ס	ע	פ	צ	ק	ר	ש	ת
L	M	N	S	O	P	Tz	Q	R	Sh	Th
12	13	14	15	16	17	18	19	20	21	22

Sh(e)Sh(a)Kh = BBL = Babel = Babylon

\* \* \* \* \*

300	30	3	200	20	2	100	10	1
Sh	L	G	R	K	B	Q	I	A
600	60	6	500	50	5	400	40	4
*M	S	V	*K	N	H	Th	M	D
900	90	9	800	80	8	700	70	7
*Tz	Tz	T	*P	P	Ch	*N	O	Z

$\overline{\text{L}}$  = A, 1  
 $\overline{\text{L}}$  = D, 4

$\overline{\text{I}}$  = I, 10  
 $\overline{\text{I}}$  = M, 40

$\overline{\text{Q}}$  = Q, 100  
 $\overline{\text{Q}}$  = Th, 400

Fig. 2.

mained a mystery, because no such place was known to geographers or historians. But then it was discovered that if you write the twenty-two letters of the Hebrew alphabet in two rows, eleven in one row and eleven in the other, as in Fig 2, you set up a substitution alphabet whereby you can replace letters by those standing opposite them. For example, "Shin", is represented by "Beth" or vice versa, so that "Sheshakh" translates "Babel", which is the old name of "Babylon". Hebrew then did not have and still doesn't have vowels, they must be supplied. This is an example of what is called ATHBASH writing, that is, where Aleph, the first letter is replaced by Teth, the last letter, Beth, the second letter, by Shin, the next-to-the-last, etc. By sliding the second row of letters one letter each time there are eleven different cipher alphabets available for use. The old Talmudists went in for cryptography to a considerable extent. Incidentally, in mentioning the Bible, I will add that Daniel, who, after Joseph in Genesis, was an early interpreter of dreams and therefore one of the first psychoanalysts, was also the first cryptanalyst. I say that he was an early psychoanalyst, because you will remember that he interpreted Nebuchadnezzar's dreams. In the Bible's own words, "Nebuchadnezzar dreamed dreams, wherewith his spirit was troubled, and sleep brake from him." But, unfortunately, when he woke up he just couldn't remember those troublesome dreams. One morning he called for his wise men, magicians, astrologers, and Chaldean sorcerers and asked them to interpret the dream he'd had during the preceding night. "Well, now, tell us the dream and we'll try to interpret it", they said. To which King Nebuchadnezzar exclaimed, "The thing is gone from me. I don't remember it. But it's part of your job to find that out, too, and interpret it. And if you can't tell me what the dream was, and interpret it, things will happen to you." What the king asked was a pretty stiff assignment, of course, and it's no wonder they failed to make good, which irked Nebuchadnezzar no end. Kings had a nasty habit of chopping your head off in those days if you failed or made a mistake, just as certain arbitrary and cruel despots are apt to do even in modern times for more minor infractions, such as not following the Party Line. So in this case it comes as no surprise to learn that Nebuchadnezzar passed the word along to destroy *all* the wise men of Babylon, among whom was one of the wise men of Israel, named Daniel. Well, when the King's guard came to fetch him, Daniel begged that he be given just a bit more time. Then, by some act of divination,—the Bible simply says that the secret was revealed to Daniel in a night vision—Daniel was able to reconstruct the dream and then to interpret it. Daniel's reputation was made. Some years later, Nebuchadnezzar's son Belshazzar was giving a feast, and, during the course of the feast, in

~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

the words of the Bible, "came forth fingers of a man's hand and wrote over against the candlestick upon the plaster of the wall" The hand wrote a secret message. You can imagine the spine-chilling scene. Belshazzar was very much upset, and just as his father did, he called for his wise men, soothsayers, Chaldean sorcerers, magicians and so on, but they couldn't read the message. Apparently they couldn't even read the cipher characters! Well, Belshazzar's Queen fortunately remembered what that Israelite Daniel had done years before and she suggested that Daniel be called in as a consultant. Daniel was called in by Belshazzar and he succeeded in doing two things. He succeeded not only in *reading* the writing on the wall "MENE, MENE, TEKEL, UPHARSIN", but also he was successful in deciphering the meaning of those strange words. His interpretation "Mene" — "God hath numbered thy kingdom and finished it" "Tekel" — "Thou art weighed in the balances and found wanting"

ingenuity of the most expert scholars of the Babylonian court. Of course it cannot be denied, as Lagarde has pointed out, that the ideographic values of these four words, 'count, mina, shekel and part,' were undoubtedly signs with which any educated Babylonian was familiar ('Mittheilungen,' iv 364). If, however, we suppose that the ideograms were written close together without any division between the individual words, a style of writing we often meet with in the cuneiform inscriptions, thus



it would be just as hard to read as a rebus and would puzzle the most skillful decipherer. The difficulty would have been still more increased if the ideograms had been grouped in some unusual way, severing the natural connection of the component elements, for example, thus



If the signs had been written in this manner it would have been almost impossible to arrive at their true meaning. The first combination, SID-MA, might have some fifteen different meanings, the second group, NA-TU-U, might signify 'is fit' or 'suitable,' while the third and last, BAR-BAR, is capable of explanation in a variety of ways. Of course, as soon as one is told the meaning of the combination, the sentence at once becomes clear.

Fig. 3.

"Upharsin" — "Thy kingdom shall be divided and given to the Medes and Persians" Apparently the chap who did the handwriting on the wall knew a thing or two about cryptography, because he used what we call "variants", or different values, for in one case the last word in the secret writing on the wall is "Upharsin" and in the other it is "Peres", the commentators are a bit vague as to why there are these two versions of the word in the Bible. At any rate, Babylon was finished, just as the inscription prophesized, it died with Belshazzar

I think this curious Biblical case of the use of cryptography is interesting because I don't think anybody has really found the true meaning of the sentence in secret writing, or explained why the writing on the wall was unintelligible to all of Belshazzar's wise men. Figure 3 is supposed to give the best explanation of the enigmatical sentence that has always been considered one of the most obscure of the many difficult scriptural passages which have awakened the interest and baffled the ingenuity of scholars. You see that this savant thinks that the cuneiform ideograms were written without any division between the individual words, so that the sentence "would be just as hard to read as a rebus and would puzzle the most skillful decipherer". He goes on to say "The difficulty would have been still more increased if the ideograms had been grouped in some unusual way, severing the natural connection of the component elements. If the signs had been written in this manner it would have been almost impossible to arrive at their true meaning". But why could Daniel read and interpret the writing when his competitors couldn't? This our savant doesn't explain. Another savant offers as his explanation of the mystery the following hypothesis. That the words were written in columns, as shown, and that Daniel in solving the mystery read downwards or rather down, up, down. This explanation doesn't satisfy me any more than the other one.

Probably the earliest reliable information on the use of cryptography in connection with an alphabetic language dates from about 900 B C, Plutarch mentioning that from the time of Lycurgus there was in use among the Lacedemonians, or ancient Greeks, a device called the *scytale*. This device, which I'll explain in a moment, was definitely known to have been used in the time of Lysander, which would place it about 400 B C. This is about the time that Aeneas Tacticus wrote his large treatise on the defense of fortification, in which there is a chapter devoted specifically to cryptography. In addition to mentioning ways of physically concealing messages, a peculiar sort of cipher disk is described. Also a method of replacing words and letters by dots is mentioned.



~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

Figure 4 is a picture of the scytale, one of the earliest cipher devices history records. The scytale was a wooden cylinder of specific dimen-

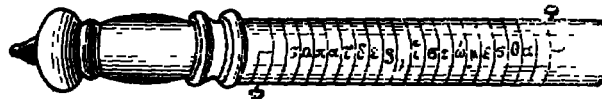
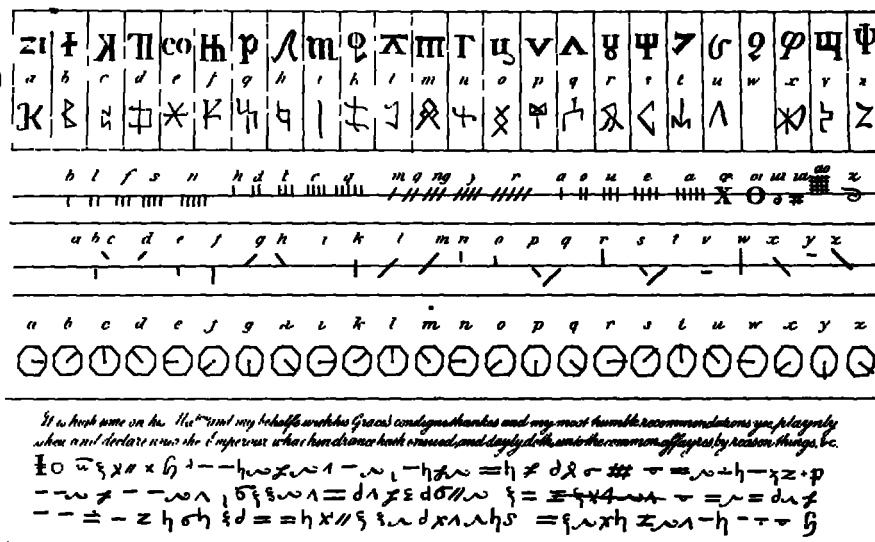


Fig. 4.

sions around which they wrapped spirally a piece of parchment or leather, they then wrote the message on the parchment, unwound it, and sent it to its destination by a safe courier, who handed it over to the commander for whom it was intended and who, having been provided with an identically-dimensioned cylinder, would wind the strip of leather or parchment around his cylinder and thus bring together properly the letters representing the message. This diagram may not be accurate. I don't think anyone really understands the scheme. The writing was done across the edges of the parchment, according to some accounts, and not between the edges, as shown here. Incidentally, you may be interested to learn that the baton which the European field marshal still carries as one of the insignia of his high office derives from this very instrument.

We don't know much about the use of cryptography by the Romans, but it is well known that Caesar used an obviously simple method, all he did was to replace each letter by the one that was fourth from it in the alphabet. For example, A would be represented by D, B by E and so on. Augustus Caesar is said to have used the same sort of thing, only even more simple: each letter was replaced by the one that followed it in the alphabet. Cicero was one of the inventors of what is now called shorthand. He had a slave by the name of Tyro, who wrote Cicero's records in what are called Tyronian notes. Modern shorthand is a development of Tyro's notation system.

In Fig 5 we see some cipher alphabets of olden times, alphabets used by certain historical figures you'll all remember. The first cipher alphabet in this figure was employed by Charlemagne, who lived from 768 to 814 A.D. The second one was used in England during the reign of Alfred the Great, 871 to 899. The third alphabet is called *ogam writing* and was used in ancient Ireland. The alphabets below that were used much later in England: the fourth one by Charles the First, in 1646, the fifth, the so-called "clock cipher", was used by the Marquis of Worcester in the 17th Century, finally, the last one was used by Cardinal Wolsey in about 1524.



**Fig. 5.**

In the Middle Ages cryptography appears first as a method of concealing proper names, usually by the simple substitution of each letter by the next one in the alphabet, just about as Augustus Caesar did hundreds of years before. At other times the vowels were replaced by dots, without changing the consonants—a method that was used throughout Europe to about 1000 A.D., when letters began to be replaced by various signs, by other letters, by letters from another language, by runes which are found in abundance in Scandinavia, and by arbitrary symbols. Figure 6 is an example of a runic inscription on a stone that stands before Gripsholm Castle near Stockholm, Sweden. The word *rune* means “secret.”

Within a couple hundred years the outlines of modern cryptography began to be formed by the secret correspondence systems employed by the small Papal States in Italy. In fact, the real beginnings of systematic, modern cryptology can be traced back to the days of the early years of the 13th Century, when the science began to be extensively employed by the princes and chanceries of the Papal States in their diplomatic relations amongst themselves and with other countries in Europe. The necessity for secret communication was first met by attempts inspired by or derived from ancient cryptography, as I've outlined so far. There was a special predilection for vowel substitution but there appeared about this time one of the

<sup>1</sup>The author's caption "A couple of old ruins"—*Ed*

~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

elements which was later to play a very prominent role in all cipher systems, an element we now call a *syllabary*, or a *repertory*. These were lists of letters, syllables, frequently-used parts of speech and words, with additions of arbitrary equivalents for the names of persons and places. There is still in existence one such syllabary and list of arbitrary equivalents which was used about 1236 A D. and there are other examples that were used in Venice in 1350



Fig. 6.

Among examples of ciphers in medieval cryptography is a collection of letters of the Archbishop of Naples, written between 1363 and 1365, in which he begins merely with symbol substitutions for the vowels and uses the letters that are actually vowels to serve as nulls or non-significant letters to throw the would-be-cryptanalyst off the right track. As a final development, the high-frequency consonants *L*, *M*, *N*, *R*, and *S*, and all the vowels, are replaced not only by arbitrary symbols but also by other letters.

~~CONFIDENTIAL~~



Cipher systems of the type I've described continued to be improved. In Figure 8 is shown what we may call the first complete cipher system of this sort. There are substitution symbols for each letter, the vowels have several equivalents, there are nulls, and there is a small list of arbitrary symbols, such as those for "the Pope", the word "and", the conjunction "with", and so on. This cipher, dated 1411, was used in Venice, and is typical of the ciphers used by the Papal chanceries of those days.

*Tavola I*

[illegible]

Dati in nostro Ducali palatio die quartiduo mensis Junij Ind. quarta millesimo centesimo

C. Insuper suis vos habere in commisso de peccatis et supplicis ut provida,  
sua per dominum papam de peccatis de quibus et quia nichil adhuc actum  
est cum declaratione et totali omnia patria et reseremus qui fiat in curia de de  
de peccatis. Mandamus fidelitatem vobis quatenus cum complicitate ad  
dominum papam, rationibus de peccatis, debetis supplicare etiam tam  
citati sua ut dignetur providere de peccatis de, et de peccatis suis peccatis  
ea qui non sit suspensus nostro dominum et qui sit gratus compatriotis  
sua cum quod de declaratione sua speamus et sunt sanctitas sua pro  
metit nam hoc etiam specialiter continetur in vestra commissione Re.  
cordando sua Beatitude quod quanto potius providetur tanto melius uti  
bus et salubris est pro patria de peccatis

A	B	C	D	E	F	G	H	I	L	M	N	O	P	Q	R	S	T	U	V	NUCLE	Q = Papa
1	p	7	8	2	5	3	1	0	-	4	9	6	5	3	5	5				b c d p	• = et
O			*			x									f		H	H		y u h	R = con
K			+			T											I	I		m o p 2	q = quo

The step remaining to be taken in the development of these ciphers was to expand the "vocabulary", that is, the list of equivalents for frequently-used words, and syllables, the names of persons and places, parts of speech, and so on. This step was reached in Italy during the

~~CONFIDENTIAL~~



first half of the 15th Century and became the prototype of diplomatic ciphers used in practically all the states of Europe for several centuries. One of 70 ciphers collected in a Vatican codex and used from about 1440 to 1469 is shown in Figure 9. Note that the equivalents of the plaintext items are Latin words and combinations of two and three letters, and that they are listed in an order that is somewhat alphabetical but not strictly so. I suppose that by constant use the cipher clerks would learn the equivalents almost by heart, so that an adherence to a strict alphabetic sequence either for the plaintext items or for their cipher equivalents didn't hamper their operations too much. In Figure 10 there is much the same sort of arrangement, except that now the cipher equivalents seem to be digraphs and these are arranged in a rather systematic order, for ease in enciphering and deciphering. Now we have the real beginnings of what we call a one-part code, that is, the same list will serve both for encoding and decoding. These systems, as I've said, remained the prototypes of the cryptography employed throughout the whole of Europe for some centuries. The Papal States used them, and as late as 1793 we find

Ar. Sim. E. Cito.

a	b	c	d	e	f	g	h	i	l	m	n
v	R	ar	w	h	e	q	p	z	d	t	&
v				a				f			
o	p	q	r	s	t	v	x	y	z		
5°	X	3	x	z	V	8°	A	Y	Ve		
6°			q	e		9°		q			
2°						10°		q			

ba	be	bi	bo	bu	ca	ce	ci	co	cu
u	v	u	u	u+	va	v	v	v	v
da	de	di	do	du	fa	fe	fi	fo	fu
u	u	u	u	u+	h	h	h	h	h
ga	ge	gi	go	gu	ha	he	hi	ho	hu
a	a	a	a	a+	o	o	o	o	o
ja	je	ji	jo	ju	la	le	li	lo	lu
e	e	e	e	e+	3+	3	3	3	3'
ma	me	mi	mo	mu	na	ne	ni	no	nu
4	4	4	4	4+	5+	5	5	5	5'
pa	pe	pi	po	pu	ga	ge	gi	go	gu
6	6	6	6	6+	n	n	n	n	n
ra	re	ri	ro	ru	sa	se	si	so	su
7	7	7	7	7+	8+	8	8	8	8'
ta	te	ti	to	tu	va	ve	vi	vo	vu
9+	9	9	9	9'	p	p	p	p	p+

Cito general 1562.

xa	xe	xi	xo	xu	ya	ye	yi	yo	yu
q+	q	q	q	q	d'	d	d	d-	d+
za	ze	zi	zo	zu	bla	ble	bli	blo	blu
f+	f	fe	f	f'	m	m	m	m-	m+
bra	bre	bri	bro	bru	cha	che	chi	cho	chu
n+	n	n	n	n	s	s	s	s-	s+
cla	cle	cli	clo	clu	cre	cre	cre	cro	cru
ψ+	ψ	ψ	ψ	ψ	φ	φ	φ	φ-	φ+
fla	fle	fli	flo	flu	fra	fre	fri	fro	fru
ω+	ω	ω	ω	ω'	g'	g	g	g-	g+
gla	gle	gli	glo	glu	gra	gre	gri	gro	gru
n+	n	n	n	n	m	m	m	m-	m+
pla	ple	pli	plo	plu	pra	pre	pri	pro	pru
N+	N	N	N	N	R	R	R	Re	R+
					tra	tre	tri	tro	tru
					D+	D	D-	D	D'

Duplices seran todas las letras, numeros  
o caracteres, que tuvieron una  
raya larguilla y llana en cima

como θ vale por xi o xi, q por z z,  
f por ss

Fig 11.



~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

them used in France I wish here to mention specifically the so-called King's General Cipher used in 1562 by the Spanish Court It is shown in Fig 11

But there were two exceptional cases which show that the rigidity of cryptographic thought was now and then broken during the four centuries we have been talking about in this brief historical survey Some of the Papal ciphers of the 16th Century and those of the French Court under Kings Louis XIII and XIV exemplify these exceptions In the case of these French Court ciphers we find that a French cryptologist named Antonio Rossignol, who was employed by Cardinal Richelieu, understood quite well the weaknesses of the one-part code and syllabaries It was he who, in about 1640, introduced a new and important improvement, the idea of the two-part code or syllabary, in which for encoding a message the items in the vocabulary are listed in some systematic order, nearly always alphabetical, the code equivalents, whatever they may be, are assigned to the alphabetically-listed items in random order This means that there must be another arrangement or book for ease in decoding, in which the code equivalents are listed in systematic order, numerically or alphabetically as the case may be, and alongside each appears its meaning in the encoding arrangement, or book The significance of this improvement you'll find out sooner or later Codes of this sort also had variants--Rossignol was clever, indeed One such code, found in the 1691 correspondence of Louis XIV had about 600 items, with code groups of two and three digits Not at all bad, for those days'

Now this sort of system would appear to be quite secure, and I suppose it was indeed so, for those early days of cryptographic development--but it wasn't proof against the cleverness of British brains, for the eminent mathematician John Wallis solved messages in it in 1689. Never underestimate the British in this science--as we'll have reason to note in another lecture in this series \*

French cryptography under Kings Louis XV and XVI declined, reaching perhaps its lowest level under Napoleon the Great It is a fact that in Napoleon's Russian enterprise the whole of his army used but a single code book of only 200 groups, practically without variants, even for the high-frequency letters Furthermore, not all the words in a message were encoded--only those which the code clerk or the writer of the message thought were important It's pretty clear that the Russians intercepted and read many of Napoleon's messages--this comes from categorical statements to this effect by Czar Alexander I

---

\* Official deciphering of foreign communications by British cryptologists can be traced back to about the year 1525, if not earlier

himself We won't be far wrong in believing that the weaknesses of Napoleon's crypto-communications formed an important factor in Napoleon's disaster A hundred and twenty-five years later, Russian ineptitude in cryptographic communications lost them the Battle of Tannenberg and knocked them out of World War I

The other 16th Century Papal ciphers that constituted the second exception to the general similarity of cryptographic systems of those days were quite different from those I've shown you In this exception the ciphers were monoalphabetic, but some letters had the same equivalent, so that on decipherment the context had to be used to decide which of two or more possible plaintext values was the one meant by each cipher letter One such cipher used by the Maltese

CIPHER OF THE INQUISITOR OF MALTA (1585)  
(From SACCO, MANUALE DI CRITTOGRAPHIA, 1947)

chiaro -	A,T	E,F	I,G	O,D	U,V,B	C,L,N	M,R	P,S,Z
cifra -	∅	3	5	4	2	6	9	7

Nulle 1,8

chiaro - qua que qui quo che chi non quando perchè,et,per

cifra -	7	9	6	2	4	5	3	∅	1,8
---------	---	---	---	---	---	---	---	---	-----

Seguono varie voci, cifrate con un gruppo di due cifre tramezzate da un punto, es 11 Papa = 2 7, 11 Re di Francia = 3 2

Fig. 12

Inquisitor in 1585 is shown in Fig 12 You'll note that the digit ∅ has two values, A and T, the digit 2 has three values, U, V, and B, and so on There were two digits used as nulls, 1 and 8, digits with dots above them stood for words such as Qua, Que, Qui, and so on

Figure 13 shows how a message would be enciphered and deciphered A bit tricky, isn't it? Many, many years later Edgar Allan Poe describes a cipher of this same general type, where the decipherer must choose between two or more possible plaintext equivalents in building up his plain text, the latter guiding the choice of the right equivalent The trouble with this sort of cipher is that you have to have pretty smart cipher clerks to operate it and even then I imagine that in many places there would be doubtful decipherments of words It wasn't really a practical system even in those days, but it could, if used skillfully and with only a small amount of text, give a cryptanalyst plenty of headaches But such systems didn't last very long because of the practical difficulties in using them

~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

## Esempio di cifratura

chiaro - S P E R O C H E O G N I C O S A S I A  
 cifra - 7 7 3 9 4 4 4 5 6 5 1 6 4 7 0 8 7 5 0  
 che sarà trasmesso tutto unito, senza spazi

## Esempio di decifrazione

cifra	4	5	1	0	2	0	4	1	4	0	9	4	8	9	5	6	2	0	4	1	0	2
	O	I		A	U	A	O		O	A	M	O		M	I	C	U	A	O		A	U
decifra	D	G		T	B	T	D		D	T	R	D		R	G	L	B	T	D		T	B
					V										N	V					V	
chiaro	D	I		T	U	T	O		D	A	R	O		M	I	N	U	T	O		A	V

cifra	5	7	4		1				4	5	6	5	1	6	4	9	5	3	9	3	8
	I	P	O		E	T			O	I	C	I		C	O	M	I	E	M	E	
decifra	G	S	D		P	E	R		D	G	L	G		N	D	R	G	F	R	F	
		Z			P	E	R	C	H		N			L							
chiaro	I	S	O		P	E	R		O	G	N	I		C	O	R	I	E	R	E	

Fig. 13.

The first regular or official cipher bureau in the Vatican was established in about 1540, and in Venice at about the same time, about one hundred years before a regular cipher bureau was established in France by Cardinal Richelieu. It is interesting to observe that no new or remarkable ideas for cryptosystems were developed for a couple of hundred years after the complex ones I've described as having been developed by the various Papal cryptologists. One-part and two-part syllabaries and simple or complex ones with variants were in use for many decades, but later on, in a few cases, the code equivalents were superenciphered, that is, the code groups formed the text for the application of a cipher, generally by rather simple systems of additives. Governmental codes were of the two-part type and were superenciphered by the more sophisticated countries.

The first book or extensive treatise on cryptography is that by a German abbot named Trithemius, who published in 1531 the first volume of a planned 4-volume monumental work. I said that he planned to publish four volumes, but he gave up after the third one, because he wrote so obscurely and made such fantastic claims that he was charged with being in league with the Devil, which was a rather dangerous association in those or even in these days. They didn't burn Trithemius but they did burn his books. Figure 14 illustrates that the necessity for secrecy in this business was recognized from the very earliest days of cryptology and certainly by Trithemius. Here is the sort of oath that Trithemius recommended be administered to

students in the science of cryptology All of you have subscribed to a somewhat similar oath, but we now go further and back up the oath with a rather strict law You've all read it, I'm sure

**The Trithemian Oath**  
 given by  
**Johannes Trithemius**  
 in

Book II, Chapter XIV, of his "Steganographia".

I, <sup>STUDENT'S NAME</sup> ---, by the Virtue of Almighty God,  
 by the Blood of our Lord Jesus Christ,  
 by the Resurrection of the Dead and  
 the last Judgment, and by the Salvation  
 of my Soul in the Holy Catholic Faith,  
 swear to Almighty God, to the Blessed Virgin  
 Mary, to all the Saints, and to you --- <sup>TEACHER'S NAME</sup> ---,  
 that I will faithfully guard this Art of  
 Steganography all the Days of my Life.  
 I will teach it to no one without your  
 Consent and Permission. Moreover I  
 likewise swear and promise that I will  
 not use this Knowledge in Opposition  
 to God and his Commandments, nor  
 in Opposition to the Holy Roman Catholic  
 Church and its Ministers.

So may God help me, and so may he  
 save me at the last Judgment.

Fig. 14.

~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

We come now to some examples from more recent history. In Fig. 15 we see a cipher alphabet used by Mary, Queen of Scots, who reigned from 1542 to 1567 and was beheaded in 1587. In this connection it may interest you to learn that question has been raised as to whether the Queen was "framed" by means of this forged postscript (Fig. 16) in a cipher that was known to have been used by her.

A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	V	X	Y	Z	
h	j	w	m	i	z	f	a	v	p	k	a	w	x	oo	:	b	b	+	+	m	y	o	f
<p>Shelton. 27. a. n. 4. 5. 2. 1. His. o. doublets.</p> <p>and as by but due for from have f m is me my no of.</p> <p>pray so say send to the that the the with usages usages.</p> <p>use you uselo latta. roodans. lattan.</p> <p>the h of frame. the h of spene. the q of england the q of scotland. the pith of the</p> <p>27 46 63 100 28</p>																							
Subscription											Superscription												
By											D												

Fig. 15.

*Handwritten cipher text, likely the forged postscript mentioned in the text.*

THE FORGED POSTSCRIPT, WITH PHILLIPS'S ENDORSEMENT  
Public Record Office

*Handwritten note:*  
 42  
 The postscript  
 of the letter  
 was written  
 by Mary

Fig. 16.

The Spanish Court under Phillip II, in the years 1555-1598, used a great many ciphers and here's one of them (Fig 17) You see that it is quite complex for those early days and yet ciphers of this sort were solved by an eminent French mathematician named Vieta, the father of modern algebra In 1589 he became a Counselor of Parliament at Tours and then Privy Counselor While in that job he solved a Spanish cipher system using more than 500 characters, so that all the Spanish dispatches falling into French hands were easily read Phillip

Cifra general 1572

xa	xe	xi	xo	xu	ya	ye	yi	yo	yu
ſ	ſ+	ſ'	ſo	ſe	h'	h+	h'	ho	hy
za	ze	zi	zo	zu	bla	ble	bli	blo	blu
i	+	i	+	+	n'	n'	n'	n'	n'
bra	bre	bri	bro	bru	cha	che	chi	cho	chu
ſ	ſ+	ſ'	ſo	ſe	d'	d'	d'	do	de
cla	cle	cli	clo	clu	cra	cre	cri	cro	cru
g'	g+	g'	go	ge	g'	g+	g'	go	ge
dra	dre	dri	dru	dru	fla	fle	fli	flo	flu
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe
fra	fre	fri	fro	fru	gla	gle	gli	glo	glu
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe
gra	gre	gri	gro	gru	pla	ple	pli	plo	plu
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe
pra	pre	pri	pro	pru	tra	tre	tri	tro	tru
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe
al	el	il	ol	ul	an	en	in	on	un
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe
ar	er	ir	or	ur	as	es	is	os	us
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe
bar	ber	bri	bro	bru	bas	bes	bis	bos	bis
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe
car	cer	cir	cor	cru	cas	ces	cis	cos	cus
ſ	ſ+	ſ'	ſo	ſe	ſ	ſ+	ſ'	ſo	ſe

Fig. 17.

was so convinced of the security of his ciphers that when he found the French were aware of the contents of his cipher dispatches to the Netherlands, he complained to the Pope that the French were using sorcery against him Vieta was called on the carpet and forced to explain how he'd solved the ciphers in order to avoid being convicted of sorcery, a serious offense.

The next cryptologist I want you to know something about is another Italian savant who wrote a book, published in 1563, in which

~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

he showed certain types of cipher alphabets that have come down in history and are famous as Porta's Alphabets. Figure 18 is an example of the Porta Table, showing one alphabet with key letters A or B, another alphabet with key letters C or D, and so on. I don't want to go into exactly how the key letters are used, it is sufficient to say that even to this day cryptograms using the Porta alphabets are occasionally encountered.

That Porta's table was actually used in official correspondence is shown by Fig. 19, which is a picture of a table found among the state

AB	a	b	c	d	e	f	g	h	i	l	m
	n	o	p	q	r	s	t	v	x	y	z
CD	a	b	c	d	e	f	g	h	i	l	m
	z	n	o	p	q	r	s	t	v	x	y
EF	a	b	c	d	e	f	g	h	i	l	m
	y	z	n	o	p	q	r	s	t	v	x
GH	a	b	c	d	e	f	g	h	i	l	m
	x	y	z	n	o	p	q	r	s	t	v
IL	a	b	c	d	e	f	g	h	i	l	m
	v	x	y	z	n	o	p	q	r	s	t
MN	a	b	c	d	e	f	g	h	i	l	m
	t	v	x	y	z	n	o	p	q	r	s
OP	a	b	c	d	e	f	g	h	i	l	m
	s	t	v	x	y	z	n	o	p	q	r
QR	a	b	c	d	e	f	g	h	i	l	m
	r	s	t	v	x	y	z	n	o	p	q
ST	a	b	c	d	e	f	g	h	i	l	m
	q	r	s	t	v	x	y	z	n	o	p
VX	a	b	c	d	e	f	g	h	i	l	m
	p	q	r	s	t	v	x	y	z	n	o
YZ	a	b	c	d	e	f	g	h	i	l	m
	o	p	q	r	s	t	v	x	y	z	n

Fig. 18.

papers of Queen Elizabeth's time, it was used for communicating with the English Ambassador to Spain. Porta was, in my opinion, the greatest of the old writers on cryptology. I also think he was one of the early but by no means the first cryptanalyst able to solve a system of keyed substitution, that is, where the key is changing consistently as the message undergoes encipherment. Incidentally, Porta also was the inventor of the photographic camera, the progenitor of which was known as the *camera obscura*.

Figure 20 is a picture of what cryptographers usually call the Vigenère Square, the Vigenère Table, or the Vigenère Tableau. It consists of a set of twenty-six alphabets successively displaced one letter per row, with the plaintext letters at the top of the square, the key-letters at the side, and the cipher letters inside. The method of using the table is to agree upon a key word, which causes the equivalents of the plaintext letters to change as the key changes. Vigenère is commonly credited with having invented that square and cipher but he really didn't and, what's more, never said he did. His table as it appears in his book, the first edition of which was published in 1586, is shown in Fig. 21. It is more complicated than as described in ordinary books on cryptology.

A·B	a	b	c	d	e	f	g	h	i	k	l	m
C·D	b	c	d	e	f	g	h	i	k	l	m	n
E·F	p	q	r	s	t	u	v	w	x	y	z	a
G·H	e	f	g	h	i	k	l	m	n	o	p	q
I·K	r	s	t	u	v	w	x	y	z	a	b	c
L·M	i	k	l	m	n	o	p	q	r	s	t	u
N·O	v	w	x	y	z	a	b	c	d	e	f	g
P·Q	n	o	p	q	r	s	t	u	v	w	x	y
R·S	u	v	w	x	y	z	a	b	c	d	e	f
T·V	o	p	q	r	s	t	u	v	w	x	y	z
x·y	f	g	h	i	k	l	m	n	o	p	q	r
z·A	s	t	u	v	w	x	y	z	a	b	c	d

Fig. 19.



~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

Figure 22 is one more example of another old official cipher. In it we can see the alphabets which could be slid up and down, as a means of changing the key. The "two-square cipher", or "two-alphabet cipher" shown in Fig. 23 is another of this type. It is a facsimile of a state cipher used in Charles the First's time, in 1627, for communicating with France and Flanders. It involves coordinates and I want you to notice that there are two complete alphabets inside it, intended to smooth out frequencies. The letters of the keywords OPTIMUS

	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. 20.

and DOMINUS serve as the coordinates used to represent the letters inside the square. A third old cipher, one used by George III in 1799 is shown in Fig. 24.

One writer deserving special attention as a knowledgeable cryptologist in the 17th Century, and the one with whose cipher I'll close this lecture, is Sir Francis Bacon, who invented a very useful cipher and mentioned it for the first time in his *Advancement of Learning*,

~~CONFIDENTIAL~~

published in 1604, in London. The description is so brief that I doubt whether many persons understood what he was driving at. But Bacon described it in full detail with examples, in his great book *De Augmentis Scientiarum* which was published almost 20 years later, in 1623, and which first appeared in an English translation by Gilbert Wats in 1640 under the title *The Advancement of Learning*. Bacon called his invention the *Biliteral Cipher* and it is so ingenious that I think you should be told about it so that you will all fully understand it.

		O	P	Q	R	S	T	V	X	A	B	C	D	E	F	G	H	I	L	M	N
		E	F	G	H	I	L	M	N	O	P	Q	R	S	T	V	X	A	B	C	D
O	E	a	b	c	d	e	f	g	h	i	l	m	n	o	p	q	r	t	v	x	
P	F	b	c	d	e	f	g	h	i	l	m	n	o	p	q	r	t	v	x	a	
Q	G	c	d	e	f	g	h	i	l	m	n	o	p	q	r	t	v	x	a	b	
R	H	d	e	f	g	h	i	l	m	n	o	p	q	r	t	v	x	a	b	c	
S	I	e	f	g	h	i	l	m	n	o	p	q	r	t	v	x	a	b	c	d	
T	L	f	g	h	i	l	m	n	o	p	q	r	t	v	x	a	b	c	d	e	
V	M	g	h	i	l	m	n	o	p	q	r	t	v	x	a	b	c	d	e	f	
X	N	h	i	l	m	n	o	p	q	r	t	v	x	a	b	c	d	e	f	g	
A	O	i	l	m	n	o	p	q	r	t	v	x	a	b	c	d	e	f	g	h	
B	P	l	m	n	o	p	q	r	t	v	x	a	b	c	d	e	f	g	h	i	
C	Q	m	n	o	p	q	r	t	v	x	a	b	c	d	e	f	g	h	i	l	
D	R	n	o	p	q	r	t	v	x	a	b	c	d	e	f	g	h	i	l	m	
E	S	o	p	q	r	t	v	x	a	b	c	d	e	f	g	h	i	l	m	n	
F	T	p	q	r	t	v	x	a	b	c	d	e	f	g	h	i	l	n	n	o	
G	V	q	r	t	v	x	a	b	c	d	e	f	g	h	i	l	m	n	o	p	
H	X	r	t	v	x	a	b	c	d	e	f	g	h	i	l	m	n	o	p	q	
I	A	t	v	x	a	b	c	d	e	f	g	h	i	l	n	n	o	p	q	r	
L	B	t	v	x	a	b	c	d	e	f	g	h	i	l	m	n	o	p	q	r	t
M	C	v	x	a	b	c	d	e	f	g	h	i	l	m	n	o	p	q	r	t	v
N	D	x	a	b	c	d	e	f	g	h	i	l	m	n	o	p	q	r	t	v	

Fig. 21.

In his *De Augmentis* Bacon writes briefly about ciphers in general and says that the virtues required in them are three "that they be

~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

easy and not laborious to write; that they be safe, and impossible to be deciphered without the key; and lastly, that they be, if possible, such as not to raise suspicion or to elude inquiry." He then goes on to say: "But for avoiding suspicion altogether, I will add another contrivance, which I devised myself when I was at Paris in my early youth, and which I still think worthy of preservation." Mind you, this was 40 years later! Let's consult Bacon for further details. In Fig. 25 we see a couple of pages of the Gilbert Wats' translation of Bacon's *De Augmentis Scientiarum*. Bacon shows what he calls "An Example of a Bi-literarie Alphabet", that is, one composed of two

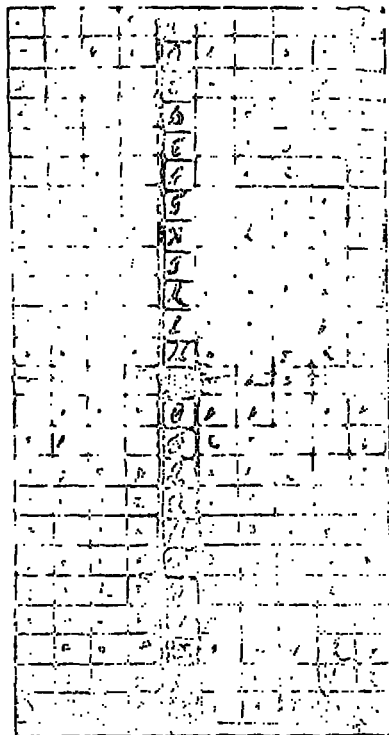


Fig. 22.



Fig. 23.

elements, which, taken in groupings of fives, yields 32 permutations. You can use these permutations to represent the letters of the alphabet, says Bacon, but you need only 24 of them [because *I* and *J*, *U* and *V*, were then used interchangeably]. These permutations of two different things—they may be "a's" and "b's", "1's" and "2's", pluses and minuses, apples and oranges, anything you please—can be used to express or signify messages. Bacon was, in fact, the inventor of the binary code which forms the basis of modern electronic digital

W. F. FRIEDMAN

—CONFIDENTIAL—

computers. Bacon gives a brief example in the word "FUGE"—the Latin equivalent for our modern "SCRAM"—as can be seen in Fig. 25. Figure 26 is another example, which quite obviously isn't what it appears to be—a crude picture of a castle, in which there are shaded and unshaded stones. It was drawn by a friend who was a physician and the message conveyed by it is:

My business is to write prescriptions  
And then to see my doses taken;  
But now I find I spend my time  
Endeavoring to out-Bacon Bacon.

* 1	A	B	C	D	E	F	G	H	I	K	L
58.	a	d	h	j	k	m	n	p	q	r	s
17.	t	v	w	x	y	z	aa	ab	ac	ad	ae
21.	af	ag	ah	ai	aj	ak	al	am	an	ao	ap
* 2	A	B	C	D	E	F	G	H	I	K	L
20.	b	c	e	f	g	i	j	k	m	n	p
87.	q	r	s	t	v	w	x	y	z	aa	ab

1. The following marks may sometimes be used  
 to signify what is intended, as in the case of  
 writing the word "Fuge" in the above. The mark & may  
 denote the whole word "Fuge" in the above.

1. The following	12. Key	11. John's
figures are to be used	13. John	16. John's
for the words which	14. John	17. John's
they are affected	15. John	18. John's
approach	16. John	19. John's

Fig. 24.

So far all this is simple enough—too much so, Bacon says, for the example he used in the case of the word FUGE is patently cryptic and would not avoid suspicion under examination. So Bacon goes on to describe the next step, which is to have at hand a “Bi-formed Alphabet”, that is, one in which all the letters of the alphabet, both capital and small, are represented by two slightly different forms of letters (Fig 28). Having these two different forms at hand, when you want to encipher your secret message you write another external and innocuous message five times as long as your secret message, using the appropriate two forms of letters to correspond to the “a’s” and “b’s” repre-

### *An Example of a Bi-literarie Alphabet*

A B C D E F  
 aaba aabab aabba aabbaaaba. aabab  
 G H I K L M  
 aabba aabbb abbaa abbaab ababa ababb  
 N O P Q R S  
 abbaa abbab abbbb abbbb baana baaba  
 T U V W X Y Z  
 baaba baabb babaa babab babba babbb

Nesher is a small matter these *Cyber-Cherubs* have, and may perform. For by this Art a way is opened, whereby a man may expresse and signifie the intentions of his munde, at any distance of place, by objects which may be preferred to the eye, and accommodated to the eare provided those objects be capable of a twofold difference onely as by Bells by Trumpets, by Lighes and Torches, by the report of Musikes, and any instruments of like nature. But to pursue our enterprife when you address your selfe to write, refigure your inward-unfolded Letter into this *Bi-litterate Alphabet*. See the *introduction Letter* be

Fugate

### Example of Solutions:

F. V. G. F  
 Абаб. бабб. абба. аабба.

**Together**

**Fig. 25.**

\* Photo, taken about December 1917, of one of several classes of student officers detailed by the Adjutant General of the U S Army to pursue a 6-weeks' course in cryptology conducted at the Riverbank Laboratories, Geneva, Illinois. Key to the cipher officers facing directly forward are "a's," officers facing either to left or right are "b's." Begin with first officer in rear row at extreme left *abaa* = *K*, *abba* = *N*, etc. Civilians seated: Colonel Fabyan, (head of the Riverbank Laboratories) at left, Mr. Friedman (Director of School) at right, Mrs. Friedman, the lady in center, other two ladies, secretaries to the Friedmans —*Ed*

senting your secret message Here's FUGE (Fig 29), enciphered within an external message saying "Manere te volo donec venero", meaning "Stay where you are until I come " In other words, whereas the real message says "SCRAM", the phoney one says "Stick around awhile, wait for me" Bacon gives a much longer example, the SPARTAN DISPATCH, here it is, and here's the secret message which it contains (Fig 30)



Fig. 26

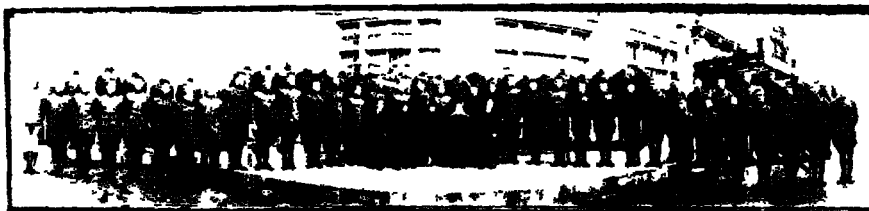


Fig 27.

Bacon's biliteral cipher is an extremely ingenious contrivance. There can be no question whatsoever about its authenticity and utility as a valid cipher. Thousands of people have checked his long example and they all find the same answer—the one that Bacon gives

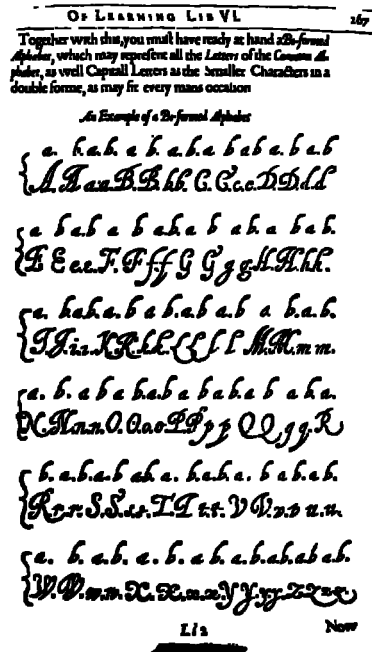
~~CONFIDENTIAL~~ INTRODUCTION TO CRYPTOLOGY

Fig. 28.

Figure 31 is a modern example which uses two slightly different fonts of type called Garamond and Imprint, and which are so nearly alike that it takes good eyes to differentiate them

The fact that Bacon invented this cipher and described it in such detail lends plausibility to a theory entertained by many persons that Bacon wrote the Shakespeare Plays and that he inserted secret messages in those plays by using his cipher. If you'd like to learn more about this theory I suggest with some diffidence that you read a book entitled *The Shakespearean Ciphers Examined*. I use the word diffidence because my wife and I wrote the book which was published in late 1957 by the Cambridge University Press.

In the next lecture we'll take up cryptology as used during the period of the American Revolution by both the Colonial and the British Forces in America.

~~CONFIDENTIAL~~

168

## OF THE ADVANCEMENT

Now to the answer letter, which is likewise you shall fit a balanced answer letter, which shall answer the other letter for letter, and afterwards fix it down. Let the answer example be,

*Mare in vobis, domine carissime.*

*An Example of Accommodation.*

*Febat V baa b baa baa.*  
*Maner te vobis donec tenore.*

We have annexed likewise a more ample example of the cypher of writing once per game. An answer letter, which is capable, we have made choice of a Spanish letter that once in a degree or second cypher'd itself.

*Ordino Rex. Rindens cariss. Rindens*  
*carissim. Regis hinc nos carissim. Regis*  
*hinc hinc maner puerum.*

An answer letter, taken out of the first Epistle of Cato, wherein a Spanish Letter is involved.

PLATE No. 50

Fig. 29.

## OF LEARNING LIBVI

*Pro omni officio, ac potius pietate erga te.*  
*caeteris satisfacio omnibus. Mihi ipse*  
*quoniam satisfacio. Tante est enim magna*  
*tudo tuorum erga me meritorum, ut quomodo*  
*am tu, nisi perfectam re, de me non conquis*  
*ci; ego, quia non idem in tua causa officio,*  
*vitam mihi esse acerbam putem. In cau*  
*sa haec sunt: Ammonius Regis legatus*  
*agere pecunia nos oppugnat. Res agitur*  
*per eosdem creditores, per quos cum tu ade*  
*nos, agebatur. Regis causa, si qui sunt,*  
*qui volunt, qui pauci sunt omnes ad longe*  
*rum rem deferri volunt. Senatus Reli*  
*gione calumniam, non religione, sed ma*  
*levolentia, et illius Regiae largitionis*  
*invidia comprobant. &c.*

PLATE No. 51

Fig. 30.

*In all duty or rather piety towards you I satisfy every body except myself. Myself I never satisfy. For so great are the services which you have rendered me, that seeing you did not rest in your endeavours on my behalf till the thing was done, I feel as if life had lost all its sweetness, because I cannot do as much in this cause of yours. The occasions are these: Ammonius the King's ambassador openly besieges us with money. the business is carried on through the same creditors who were employed in it when you were here, &c.*

Fig. 31.



~~CONFIDENTIAL~~

## Introduction to Cryptology—III

BY WILLIAM F. FRIEDMAN

*Confidential*

*The third lecture in this series deals with the cryptosystems employed by the British Regulars and by the Colonials during the period of the American Revolution. This is followed by a brief explanation of the cryptanalytic nature of the initial breaks in the solution of the ancient Egyptian hieroglyphic writing.*

Continuing with our survey of cryptologic history, the period of the American Revolution, in U S history, is naturally of considerable interest to us and warrants more than cursory treatment. Information regarding the codes and ciphers employed during that period has been rather sparse until quite recently, when a book entitled *Turncoats, Traitors and Heroes* by Col John Bakeless, AUS, was published in 1959 by Lippincott. After a good many years of research Col Bakeless brought together for the first time a considerable amount of authentic information on the subject, and some of it is incorporated in this lecture.

According to Col Bakeless—and believe it or not—in early 1775 the British commander-in-chief in America, General Gage, had no code or cipher at all, nor even a staff officer who knew how to compile or devise one, he had to appeal to the commanding general in Canada, from whom he probably obtained the single substitution cipher which was used in 1776 by a British secret agent who—again, believe it or not—was General Washington's own director-general of hospitals, Dr Benjamin Church. General Washington had means for secret communication from the very beginning of hostilities, probably even before the fighting began at Lexington and Concord. If the British under General Gage were poorly provided in this respect, by the time Sir Henry Clinton took over from General Howe, who succeeded Gage, they were much better off—they had adequate or apparently adequate means for secret communication.

Are you astonished to learn that the systems used by the American colonial forces and by the British regulars were almost identical? You shouldn't be, because the language and backgrounds of both were identical. In one case, in fact, they used the same dictionary as a code book, something which was almost inevitable because there were so few English dictionaries available. Here's a list of the systems they used.

~~CONFIDENTIAL~~

a Simple, monoalphabetic substitution—easy to use and to change

*b* Monoalphabetic substitution with variants, by the use of a long key sentence I'll show you presently an interesting example in Benjamin Franklin's system of correspondence with the elder Dumas

### c The Vigenère cipher with repeating key

#### **d Transposition ciphers of simple sorts**

<sup>e</sup> Dictionaries employed as codebooks, with and without added encipherment Two were specially favored, Entick's *New Spelling Dictionary*, and Bailey's *English Dictionary* A couple of pages from

178 J A C J A U

**Hyp**, *v. a.* to make melancholy, to dispirit  
**Hypa/lis**, *f.* a change of tastes, &c.  
**Hyperbole**, *f.* an exaggeration, a diminution  
**Hyperbolic**, *a.* exaggerating or extenuating  
**Hyperborean**, *a.* northern (reson)  
**Hyper**, **Hyperic**, *f.* a critic exact beyond  
**Hypercritical**, *a.* critical beyond use, severe  
**Hypermeter**, *f.* what is above the standard  
**Hyperbatics**, *f.* a growth of proud flesh  
**Hyphen**, *f.* ( ) between words or syllables  
**Hyponic**, *f.* a med cane causing sleep  
**Hyperchondriac**, *f.* one affected with melancholy  
**Hyperchondriacal**, *a.* melancholy  
**Hypocit**, *f.* dissimulation, a pretence  
**Hypocrite**, *f.* a dissembler in religion, &c.  
**Hypocritical**, *a.* dissembling, insincere, false  
**Hypocritically**, *ad.* without sincerity, false y  
**Hypogastria**, *a.* in the lower part of the belly  
**Hypocrit**, *f.* a distinct substance, personality  
**Hypocritical**, *a.* constitutive, distinct, personal  
**Hypothet**, *f.* a system upon supposition  
**Hypothetical**, *a.* supposed, conditional  
**Hypothesis**, *ad.* upon supposition  
**Hyst**, **Hurd** or **Herk**, *f.* a wood  
**Hyst**, *p.* a plant  
**H**, *a.* a troubled with fits  
**Hysteria**, *f.* fits of women  
1.

**I**, *pron* myself  
**Jabber**, *v. a.* to talk idly, to chatter  
**Jabberer**, *f.* one who talks unintelligibly  
**Jacent**, *a.* lying at length, extended  
**Jacinth**, *f.* a gem, the hyacinth  
**Jack**, *f.* John, an engine, fish leatheren cann  
**Jackal**, *f.* a beast that harts the lion's prey  
**Jaculent**, *f.* a simple sheepish fellow  
**Jack-anape**, *f.* a monkey, a comcord  
**Jackboots**, *f.* boots leaving for armor

**Jackdaw**, *f.* a chattering bird  
**Jackets**, *f.* a waistcoat, a short coat  
**Jackpudding**, *f.* a merry andrew, a buffoon  
**Jackmate**, *f.* a partizan of James II  
**Jactitation**, *f.* a boasting motion reflectiveness  
**Jactitation**, *f.* the act of throwing or darding  
**Jade**, *f.* a bad woman, a worthless horse  
**Jade**, *v. a.* to tire, weary, ride down, sink  
**Jadish**, *a.* unruly, vicious, unchangeable  
**Jagg**, *v. a.* to notch, *f.* a denatulation, unevenness  
**Jaggles**, *f.* a cutting in notches  
**Jaggy**, *a.* uneven, notched  
**Jail**, *f.* a prison, a goal  
**Jailer**, *f.* the keeper of a prison  
**Jakes**, *f.* a house of office, a boghouse  
**Jam**, *f.* a conserve of fruit, a child's frock  
**Jam**, *v. a.* to confine between, to wedge in  
**Jamb**, *f.* the upright post of a door  
**Jambvie**, *f.* verses composed of a long and a short  
Silable alternately  
**Jangle**, *v. a.* to wrangle, to be out of tune  
**Jannary**, *f.* a Turkish soldier, a guard  
**Janty**, *a.* showy, glittering, gay, giddy  
**Jenary**, *f.* the first month of the year  
**Japan**, *f.* a varnish to work in colors  
**Japan**, *v. a.* to varnish, to blacken  
**Japanner**, *f.* a shoeblack, one who japans  
**Jar**, *v. a.* to clash, disagree, differ, quarrel  
**Jar**, *f.* discord, a harsh sound, an earthen vessel  
**Jargon**, *v. a.* to confound, perplex, pervert  
**Jargon**, *f.* gibberish, gabble, nonsense  
**Jasmine**, *f.* a flower  
**Jasper**, *f.* a precious green stone  
**Javelin**, *f.* a spear or half pike  
**Jaw-dice**, *f.* a dissembler  
**Jaw-diced**, *a.* affected with the jaundice  
**Jaw**, *v. a.* to walk or travel about  
**Jawnt**, *f.* a ramble, a scurion, a petty  
[Jawnt],

**Fig. 1.**

the former are shown in Fig 1 To represent a word by code equivalent you simply indicated the page number, then whether column 1 or column 2 contained the word you wanted, and then the number of the word in the column Thus The word "jacket" would be represented by 178-2-2

**f. Small, specially-compiled, alphabetic one-part codes of 600-700**

items and code names—our old friend the syllabary, or repertory, of hoary old age, but in new dress In some cases these were of the “one-part” or “alphabetic” type

*g* Ordinary books, such as Blackstone's *Commentaries on the Laws of England*, giving the page number, the line number and the letter number in the line, to build up, letter-by-letter, the word to be represented Thus 125-12-17 would indicate the 17th letter in the 12th line on page 125, it might be the letter T

*h* Secret inks Both the British and the Americans made extensive use of this method

*i* Special designs or geometric figures, such as one I'll show you presently

*j* Various concealment methods, such as using hollow quills of large feathers or hollowing out a bullet and inserting messages written on very thin paper Strictly speaking, however, this sort of stratagem doesn't belong to the field of cryptology But it's a good dodge, to be used in special cases

In the way of ciphers a bit more complex than simple monoalphabetic substitution ciphers, the British under Clinton's command used a system described by Bakeless in the following terms

“ a substitution cipher in which the alphabet was reversed, ‘z’ becoming ‘a’ and ‘a’ becoming ‘z’ To destroy frequency clues, the cipher changed in each line of the message, using ‘y’ for ‘a’ in the second line, ‘x’ for ‘a’ in the third, and so on When the cipher clerk reached ‘o’ in the middle of the alphabet, he started over again A spy using this cipher did not have to carry incriminating papers, since the system was so easy to remember ”

The alphabets of this scheme are simple reversed standard sequences

ABCDEF GHIKLMNOPQRSTUWXYZ  
ZYXWUTSRQPONMLKIHGFEDCBA  
YXWUTSRQPONMLKIHGFEDCBAZ  
XWUTSRQPONMLKIHGFEDCBAZY

ONMLKIHGFEDCBAZYXWUTSRQP

Bakeless doesn't explain why the cipher sequences are only 12 in number—nor does the source from which he obtained the information, a note found among the *Clinton Papers* in the Clements Library at the University of Michigan.

Bakeless continues

“Clinton also used another substitution cipher, with different alphabets for the first, second and third paragraphs Even if an American cryptanalyst should break the cipher in one paragraph, he would have

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

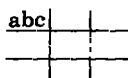
to start all over in the next As late as 1781, however, Sir Henry was using one extremely clumsy substitution cipher, in which 'a' was 51, 'd' was 54, 'e', 55 Finding that 'a' was 51 and 'd' was 54, anyone could guess (correctly) that 'b' was 52, 'c' 53 Somewhat more complex was his 'pigpen' cipher, in which twenty-five letters of the alphabet were placed in squares Then an angle alone would represent a letter, the same angle with a dot another letter, the same angle with two dots still another In some cases, cryptography was used only for a few crucial words in an otherwise 'clear' message, a method also favored by certain American officials "

Of the first cipher mentioned in the preceding extract, there is much more to be said Perhaps Bakeless was limited by space considerations In any case, I will leave that story for another time and place As for the second cipher Bakeless mentions in the extract, I can give you the whole alphabet, for it exists among the *Clinton Papers*

A B C D E F G H I K L M N O P Q R S T U W X Y Z  
51 52 53 54 55 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78

There is no explanation why the sequence beginning with 50 stops with E-55 and then, starting with F-60 goes straight on without any break to Z-78 (Remember that in those days I and J were used interchangeably, as were U and V)

Finally, as to what Bakeless (and others) call the "pigpen" cipher, this is nothing but the hoary old so-called "Masonic" cipher based upon the 4-cross figure



a    ┘    b- ┘    c- ┘

which can accommodate 27 characters, not 25, as Bakeless indicates Letters can be inserted in the design in many different arrangements

I've mentioned that code or conventional names were used to represent the names of important persons and places in these American colonial and British cryptograms of the Revolution Here are examples selected from a list of code names prepared by the famous British spy, Major André, chief of intelligence under General Clinton.

For American Generals--the names of the Apostles, for instance

General Washington was *James*

General Sullivan was *Matthew*

Names of Forts

Fort Wyoming - *Sodom*

Fort Pitt - *Gomorrha*

Names of Cities

Philadelphia - *Jerusalem*

Detroit - *Alexandria*

W. F. FRIEDMAN

~~CONFIDENTIAL~~

## Names of Rivers and Bays:

Susquehanna — *Jordan*Delaware — *Red Sea*

## Miscellaneous.

Indians — *Pharisees*Congress — *Synagogue*

I'm sure you've learned as school children all about the treasonable conduct of Benedict Arnold when he was in command of the American Forces at West Point; but you probably don't know that practically all his exchanges of communications with Sir Henry Clinton, Commander of the British Forces in America, were in cipher, or in invisible inks. One of Arnold's cipher messages, in which he offers to give up West Point for £20,000, is shown in Fig. 2; Fig. 2a

Fig. 2a.

Fig. 2b.

being the secret version, Fig. 2b, the plain text. Arnold left a few words *en clair*, the ones he considered unimportant; for the important ones he used a dictionary as a codebook, indicating the page number, column number and line number corresponding to the position in the dictionary of the plaintext word which the code group represents. Arnold added 7 to these numbers, which accounts for the fact that the first number in a code group is never less than 8, the central number is always either 8 or 9, and the third number is never less than 8 or more than 36. The significant sentence appears near the middle of the message: "If I 198-9-34, 185-8-31 a 197-8-8 . . ." yields the plain text: If I point out a plan of cooperation by which S. H. (Sir Henry Clinton) shall possess himself of West Point, the Garrison, etc., etc., etc., twenty thousands pound Sterling I think will be a cheap purchase for an object of so much importance." The signa-

110 g 14 286.9 g 93 — Tim 190 g 94.9 g 62  
 at 175 g 11 116 8.77 285.9 g 94 morning 172.8  
 on 171 290 287.6 93 at 118 g 26 169 g 60  
 289 g 16 6.11 167 g 27 171.16 g 23  
 12 g 17 and 120 g 18 and 290.9 g 27  
 160 g 143 at 190 g 32 157 g 20

General Hefington will be at Long Point Sunday,  
leaving next on his way to the Island where he is to  
be French Admiral and I am at hand will help at P. M. 11.15.

**Fig. 3.**

Sir

W Howe

is gone to the

Chesapeake bay with  
the greatest part of the  
army I hear he is now  
landed but am not  
certain I am  
left to command  
here with a  
too small force

too small force

to make any effectual  
diversion in your favor.

I shall try something cer

At any rate It may be of use  
to you I own to you I think  
S<sup>r</sup> W's move just at this time  
the worst he could take  
much joy on your success

[illegible][illegible]

**Fig. 4.**

~~CONFIDENTIAL~~

tions with him. The real or significant text is written in lines outlined by an hour-glass figure and then dummy words are supplied to fill up the lines so that the entire letter apparently makes good sense. To read the secret message, you're supposed to have the same size hour-glass figure that was used to conceal the secret message. In Fig 4 the left-hand portion shows the "phoney" message. Masks having small rectangular apertures were also used, the significant words being written so that they were disclosed when the mask was placed on the written message so as to isolate them from the non-significant words. The significant text in this example is shown

in printed form to the right of the original hour-glass design



Fig. 5.

Arnold even used the trick, mentioned above in method J, that was quite similar to one used recently by the Russian spy, Colonel Abel (Fig. 5) who was arrested in New York in June 1957, tried and convicted, and is still languishing in a Federal prison.

An interesting episode involving concealment of this sort is recorded by Bakeless. An urgent message from Sir Henry Clinton, dated 8 October 1777, and written on thin silk, was concealed in an oval silver ball, about the size of a rifle bullet, which was handed to Daniel Taylor, a young officer who had been promised promotion if he got through alive. The bullet was made of silver, so that the spy could swallow it without injury from corrosion. Almost as soon as he started, Taylor was captured. Realizing his peril too late, the spy fell into a paroxysm of terror and, crying, "I am lost," swallowed the silver bullet. Administration of a strong emetic soon produced the bullet with fatal results, for Taylor was executed. "A rather heartless American joke went around," adds Bakeless, "that Taylor had been condemned 'out of his own mouth'."

We next see (Fig. 6) one Benedict Arnold message that never was deciphered. It is often referred to as "Benedict Arnold's Treasonable Cow Letter." Only one example is extant, certain words have purely arbitrary meanings, as prearranged. The letter was written just two weeks before the capture of Major André.

In Fig 7, we see a British cipher message of the vintage 1781. It was deciphered *before* finding the key, always a neat trick when or if you can do it. The key—the title page of the then current British

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

Handwritten note: *Handwritten note: Mar 18 1896*

Dear Sir

I have bought a new and better  
 than the other people. I have also lost the. At the  
 last of the year, he is now gone to the times with  
 a very bad report to return and will return the  
 to the last of the year on Monday. He is  
 for a while. I am to give him the present on 18.12  
 and I am sure of good luck. I have and the  
 strong to you & by you will be good enough  
 to determine my 18.12. I will be in the  
 last, and had the a before hand with the same  
 about to me. The day is now gone with the  
 very much of the past.

I am  
 27.

your old friend

Alfred

Handwritten note: *Handwritten note: 18.12*

Fig. 6.

No 10 - (1701) - Dec 13

Fig. 7.



Army List -is shown in Fig 8 The numbers in the cipher text obviously refer to line numbers and letter numbers in the line of a key text, the first series of numbers, viz., 22 6 7 39 5 9 17, indicating line number 22, letter numbers 6 7 39 5 9 17 in that line Because of so many repetitions, the plain text was obtained by straightforward analysis by an officer recently on duty in NSA, Captain Edward W Knepper, USN, to whom I am indebted for this interesting example The plain text, once obtained, gave him clues as to what the key text might be, simply by placing the plaintext letters in their numerical-equivalent order in the putative key text This done, Captain Knepper was quick to realize what the key text was—a British Army List The date of the message enabled him to find the list without much difficulty in the Library of Congress (Fig 8)

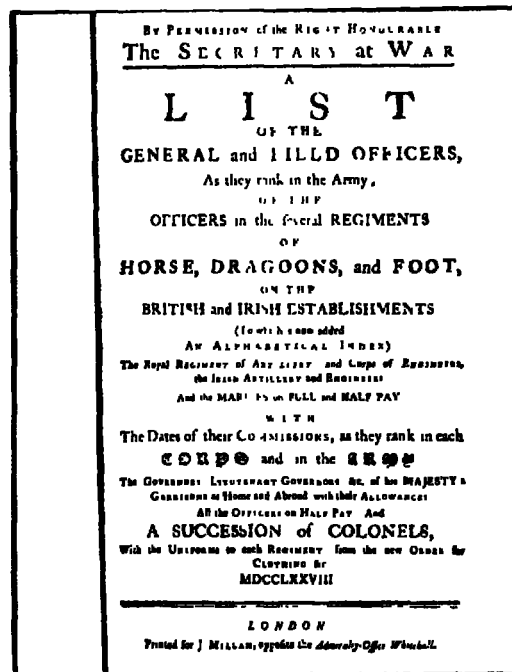


Fig. 8

There was an American who seems to have been the Revolution's one-man National Security Agency, for he was the one and only cryptologic expert Congress had, and, it is claimed, he managed to decipher nearly all, if not all, of the British code messages obtained in one way or another by the Americans Of course, the chief way in which enemy messages could be obtained in those days was to capture couriers, knock them out or knock them off, and take the

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

messages from them This was very rough stuff, compared to getting the material by radio intercept, as we do nowadays

I think you'll be interested to hear a bit more about that one-man NSA His name was James Lovell and besides being a self-trained cryptologist, he was also a member of the Continental Congress. There's on record a very interesting letter which he wrote to General Nathaniel Greene, with a copy to General Washington Here it is

Philadelphia, Sept 21, 1781

Sir

You once sent some papers to Congress which no one about you could decypher Should such be the Case with some you have lately forwarded I presume that the Result of my pains, here sent, will be useful to you I took the Papers out of Congress, and I do not think it necessary to let it be known here what my success has been in the attempt For it appears to me that the Enemy make only such Changes in their Cypher, when they meet with misfortune, as makes a difference of Position only to the same Alphabet, and therefore if no talk of Discovery is made by us here or by your Family, you may be in Chance to draw Benefit this Campaign from my last Night's Watching

I am Sir with much respect,

Your Friend,  
JAMES LOVELL

Maj Genl Greene  
(With copy to Genl Washington)

In telling you about Lovell I should add to my account of that interesting era in cryptologic history an episode I learned about only recently When a certain message of one of the generals in command of a rather large force of Colonials came into Clinton's possession he sent it off post haste to London for solution Of course, Clinton knew it was going to take a lot of time for the message to get to London, be solved and returned to America—and he was naturally a bit impatient He felt he couldn't afford to wait that long Now it happened that in his command there were a couple of officers who fancied themselves to be cryptologists and they undertook to solve the message, a copy of which had been made before sending the original off to London Well, they gave Sir Henry their solution and he acted upon it The operation turned out to be a dismal failure, because the solution of the would-be-cryptanalysts happened to be quite wrong! The record doesn't say what Clinton did to those two unfortunate cryptologists when the correct solution arrived from London some weeks later By the way, you may be interested in learning that the British operated a regularly-established cryptanalytic bureau as early as in the year 1630 and it continued to operate until the end of July 1844 Then there was no such establishment until World War I I wish there were time to tell

~~CONFIDENTIAL~~



~~CONFIDENTIAL~~

## HISTORY OF CRYPTOLOGY

libraries used by Thomas Jefferson. It is constructed on the so-called two-part principle, which was explained in the preceding lecture. Fig 9a is a portion of the encoding section, and Fig 9b is a portion of the decoding section, in which the code equivalents are in numerical order accompanied by their meanings as assigned them in the encoding section. This sort of system, which, as I've already explained, was quite popular in Colonial times as in the early days of Italian cryptography, is still in extensive use in some parts of the world.

A few minutes ago I mentioned Benjamin Franklin's cipher system, which, if used today, would be difficult to solve, especially if there were only a small amount of traffic in it. Let me show you what it was. Franklin took a rather lengthy passage from some book in French and numbered the letters successively. These numbers then became equivalents for the same letters in a message to be sent. Because the key passage was in good French, naturally there were many variants for the letter E—in fact, there were as many as one would expect in normal plain-text French, the same applied to the other high-frequency letters such as R, N, S, I, etc. What this means, of course, is that the high-frequency letters in the plain text of any message to be enciphered could be represented by many different numbers and a solution on the basis of frequency and repetitions would be very much hampered by the presence of many variant values for the same plaintext letter. In Fig 10 you can see this very clearly.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

Fig. 10.

I know of but one case in all our U S history in which a resolution of Congress was put out in cryptographic form. It is shown in Fig. 11—a resolution of the Revolutionary Congress dated 8 February 1782. I have in my collection not only a copy of the resolution but also a copy of the syllabary by which it can be deciphered.

Interest in cryptology in America seems to have died with the

~~CONFIDENTIAL~~

W. F. FRIEDMAN

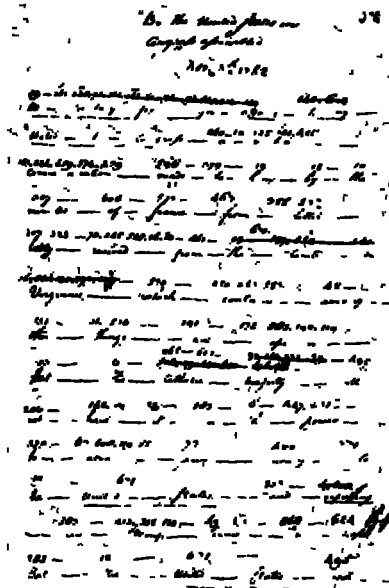
~~CONFIDENTIAL~~

Fig. 11.

passing of Jefferson and Franklin. But if interest in cryptology in America wasn't very great, if it existed at all after the Revolution, this was not the case in Europe. Books on the subject were written, not by professionals, perhaps, but by learned amateurs, and I think you will find some of them in the NSA library if you're interested in the history of the science. The next illustration (Fig. 12) is the frontispiece of a French book the title of which (translated) is "Counter-espionage, or keys for all secret communications." It was published in Paris in 1793. In the picture, we see Dr. Cryppy himself, and perhaps a breadboard model of a GS-11 research analyst, or maybe an early model of a WAC.

I am now going to tell you something about the early steps in finding an answer to the age-old mystery presented by Egyptian hieroglyphics, not only because I think that the solution represents the next landmark in the history of cryptology, but also because the story is of general interest to any aspiring cryptologist. About 1821 a Frenchman, Champollion, startled the world by beginning to publish translations of Egyptian hieroglyphics, although in the budding new field of Egyptology much had already transpired and been published. In Fig. 13 we see the gentleman and in Fig. 14, a picture of the great Napoleonic find that certainly facilitated and perhaps made possible the solution of the Egyptian hieroglyphic writing—the Rosetta Stone. The Rosetta Stone was found in 1799 at Rashid, or,

~~CONFIDENTIAL~~

## HISTORY OF CRYPTOLOGY



Fig. 12.

as the Europeans call it, Rosetta, a town in northern Egypt on the west bank of the Rosetta branch of the Nile. Rosetta was in the vicinity of Napoleon's operations which ended in disaster. When the peace treaty was written, Article 16 of it required that the Rosetta Stone, the significance of which was quickly understood by both the conquered French and victorious British commanders, be shipped to London, together with certain other large antiquities. The Rosetta Stone still occupies a prominent place in the important exhibits at the British Museum. The Rosetta Stone is a bi-lingual inscription, because it is in Egyptian and also Greek. The Egyptian portion consists of two parts, the upper one in hieroglyphic form, the lower



Fig. 13.

~~CONFIDENTIAL~~

one in a sort of cursive script, also Egyptian but called "Demotic" It was soon realized that all three texts were supposed to say the same thing, of course, and since the Greek could easily be read, it served as something called in cryptanalysis a "crib." Any time you are lucky enough to find a crib it saves you hours of work. It was by means of this bi-lingual inscription that the Egyptian hieroglyphic

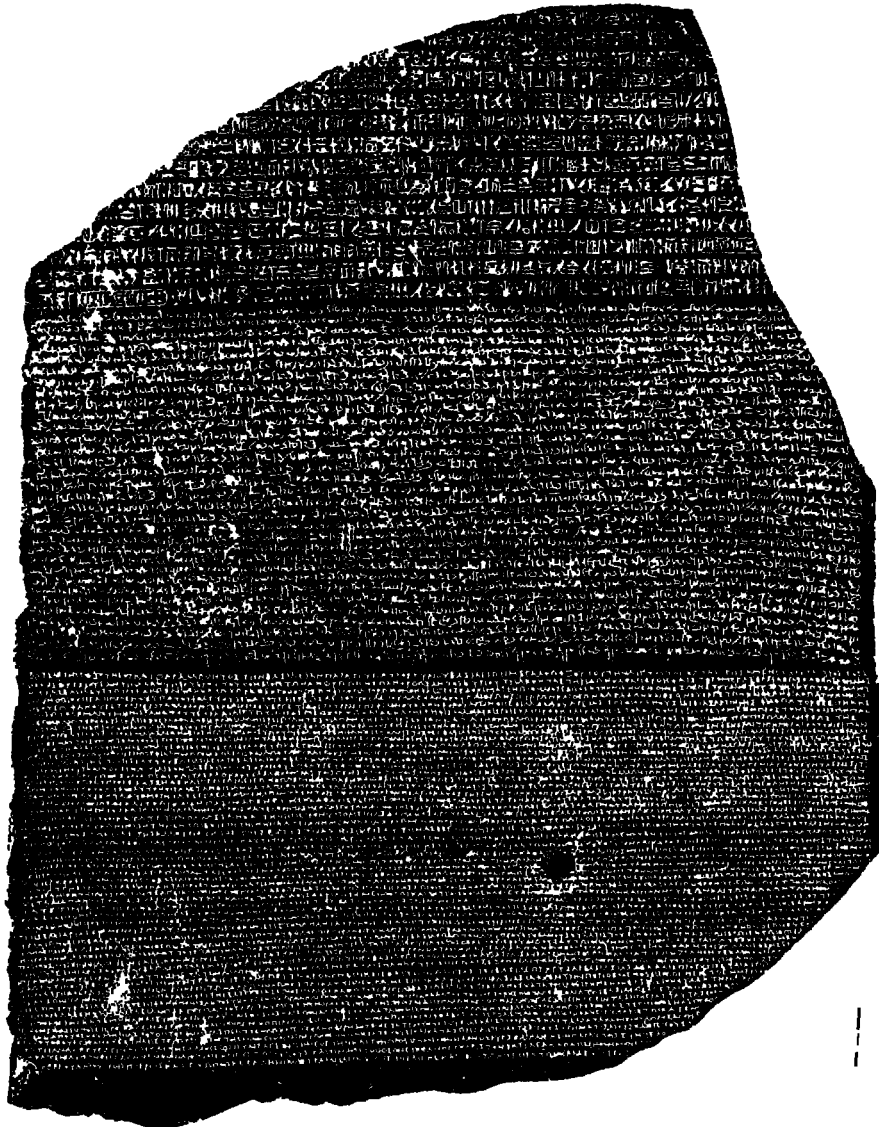


Fig. 14.

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

writing was finally solved, a feat which represented the successful solution to a problem the major part of which was linguistic in character. The cryptanalytic part of the task was relatively simple. Nevertheless, I think that anyone who aspires to become a professional cryptologist should have some idea as to what that cryptanalytic feat was, a feat which some professor (but not of cryptologic science. I think it was Professor Norbert Wiener, of the Massachusetts Institute of Technology) said was the greatest cryptanalytic feat in history. We shall see how wrong the good professor was, because I'm going to demonstrate just what the feat really amounted to by showing you some simple pictures.

First, let me remind you that the Greek text served as an excellent crib for the solution of both Egyptian texts, the hieroglyphic and the Demotic, the latter merely being the conventional abbreviated and modified form of the Hieratic character or cursive form of hieroglyphic writing that was in use in the Ptolemaic Period.

The initial step was taken by a Reverend Stephen Weston who made a translation of the Greek inscription, which he read in a paper delivered before the London Society of Antiquaries, in April 1802.


In 1818 Dr. Thomas Young, the physicist who first proposed the wave theory of light, compiled for the 4th volume of *Encyclopaedia Britannica*, published in 1819, the results of his studies on the Rosetta Stone and among them there was a list of several Egyptian characters to which, in most cases, he had assigned correct phonetic values. *He was the first to grasp the idea of a phonetic principle in the Egyptian hieroglyphs and he was the first to apply it to their decipherment.* He also proved something which others had only suspected, namely, that the hieroglyphs in ovals or cartouches were royal names. But Young's name is not associated in the public mind with the decipherment of Egyptian hieroglyphics—that of Champollion is very much so. Yet much of what Champollion did was based upon Young's work. Perhaps the greatest credit should go to Champollion for recognizing the major importance of an ancient language known as Coptic as a bridge that could lead to the decipherment of the Egyptian hieroglyphics. As a lad of seven he'd made up his mind that he'd solve the hieroglyphic writing and in the early years of the 19th Century he began to study Coptic. In his studies of the Rosetta Stone his knowledge of Coptic, a language the knowledge of which had never been lost, enabled him to deduce the phonetic value of many syllabic signs, and to assign correct readings to many pictorial characters, the meanings of which became known to him from the Greek text on the Stone.

The following step-by-step account of the solution is taken from

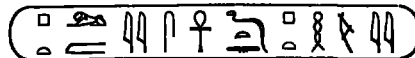
~~CONFIDENTIAL~~



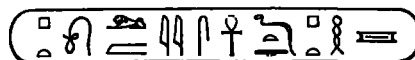
a little brochure entitled *The Rossetta Stone*, published by the Trustees of the British Museum. It was written in 1922 by E. A. Wallis Budge and was revised in 1950. I quote

"The method by which the greater part of the Egyptian alphabet was recovered is this. It was assumed correctly that oval , or "cartouche" as it is called, always contained a royal name. There is only one cartouche (repeated six times with slight modifications) on the Rosetta Stone, and this was assumed to contain the name of Ptolemy, because it was certain from the Greek text that the inscription concerned a Ptolemy. It was also assumed that if the cartouche did contain the name of Ptolemy, the characters in it would have the sounds of the Greek letters, and that all together they would represent the Greek form of the name of Ptolemy. Now on the obelisk which a certain Mr Banks had brought from Philae there was also an inscription in two languages, Egyptian and Greek. In the Greek portion of it two royal names are mentioned, that is to say, Ptolemy and Cleopatra, and on the second face of the obelisk there are two cartouches, which occur close together, and are filled with hieroglyphs which, it was assumed, formed the Egyptian equivalents of these names. When these cartouches were compared with the cartouche on the Rosetta Stone it was found that one of them contained hieroglyphic characters that were almost identical with those which filled the cartouche on the Rosetta Stone. Thus there was good reason to believe that the cartouche on the Rosetta Stone contained the name of Ptolemy written in hieroglyphic characters. The forms of the cartouches are as follows

On the Rosetta Stone —



On the Obelisk from Philae —

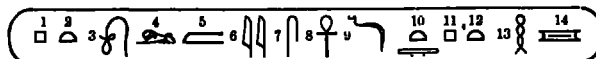


In the second of these cartouches a single sign takes the place of three signs at the end of the first cartouche. Now it has already been said that the name of Cleopatra was found in Greek on the Philae Obelisk, and the cartouche which was assumed to contain the Egyptian equivalent to this name appears in this form



Taking the cartouches which were supposed to contain the names of Ptolemy and Cleopatra from the Philae Obelisk, and numbering the signs we have

Ptolemy, A



Cleopatra, B



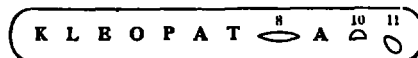
~~CONFIDENTIAL~~

## HISTORY OF CRYPTOLOGY

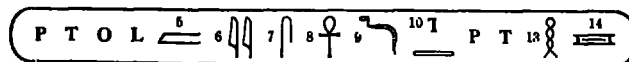
Now we see at a glance that No 1 in A and No 5 in B are identical, and judging only by their position in the names they must represent the letter P. No 4 in A and No 2 in B are identical, and arguing as before from their position, they must represent the letter L. As L is the second letter in the name of Cleopatra, sign No 1 in B must represent K. In the cartouche of Cleopatra, we now know the values of Signs Nos 1, 2 and 5, so we may write them down thus



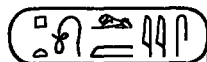
In the Greek form of the name of Cleopatra there are two vowels between the L and the P, and in the hieroglyphic form there are two hieroglyphs, this and this , so we may assume that the first is E and the other O. In some forms of the cartouche of Cleopatra, No 7 (the hand) is replaced by a half circle, which is identical with No 2 in A and No 10 in B. As T follows P in the name Ptolemy, and as there is a T in the Greek form of the name of Cleopatra, we may assume that the half circle and the hand have substantially the same sound, and that that sound is T. In the Greek form of the name Cleopatra there are two A's, the position of which agree with No 6 and No 9, and we may assume that the bird has the value of A. Substituting these values for the hieroglyphs in B we may write it thus

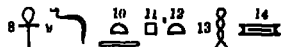



Thomas Young noticed that the two signs and always followed the name of a goddess, or queen, or princess. Other early decipherers regarded the two signs as a mere feminine termination. The only sign for which we have no phonetic equivalent is No 8, the lens, and it is obvious that this must represent R. Inserting this value in the cartouche we have the name Cleopatra deciphered. Applying now the values which we have learned from the cartouche of Cleopatra to the cartouche of Ptolemy, we may write it thus

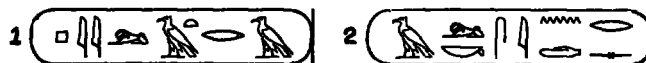


We now see that the cartouche must be that of Ptolemy, but it is also clear that there must be contained in it many other hieroglyphs which do not form part of his name. Other forms of the cartouche of Ptolemy are found, even on the stone, the simplest of them written thus

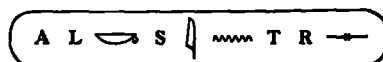
~~CONFIDENTIAL~~




It was therefore evident that these other signs  were

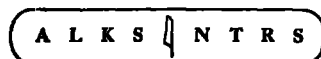
royal titles corresponding to those found in the Greek text on the Rosetta Stone meaning "ever-living, beloved of Ptah." Now the Greek form of the name Ptolemy, i.e. Ptolemaios, ends with S. We may assume therefore that the last sign<sup>1</sup> in the simplest form of the cartouche given above has the phonetic value of S. The only hieroglyphs now doubtful are , and their position in the name of Ptolemy suggests that their phonetic values must be M and some vowel sound in which the I sound predominates. These values, which were arrived at by guessing and deduction, were applied by the early decipherers to other cartouches, e.g.




Now, in No 1, we can at once write down the values of all the signs, viz, P I L A T R A, which is obviously the Greek name Philotera. In No 2 we know only some of the hieroglyphs, and we write the cartouche thus



It was known that the running-water sign  occurs in the name Berenice, and that it represents N, and that this sign  is the last word of the transcript of the Greek title "Kaisaros," and therefore represents some S sound. Some of the forms of the cartouche of Cleopatra begin with , and it is clear that its phonetic value must be K. Inserting these values in the cartouche above we have



which is clearly meant to represent the name "Alexandros," or Alexander. The position of this sign () shows that it represented some sound of E or A.

Well, I've showed you enough to make fairly clear what the problem was and how it was solved. As you may already have gathered, the cryptanalysis was of a very simple variety.

The grammar?—Well, that's an entirely different story: There's where the difficult part lay. It was very fortunate that the first attacks on Egyptian hieroglyphics didn't have to deal with enciphered writing. Yes, the Egyptians also used cryptography; yes, there *are* "cryptographic hieroglyphics!" We'll get to these later, but at this point it may be of interest to many of you to learn something about what the Rosetta Stone had to say, as set forth by Dr Budge.

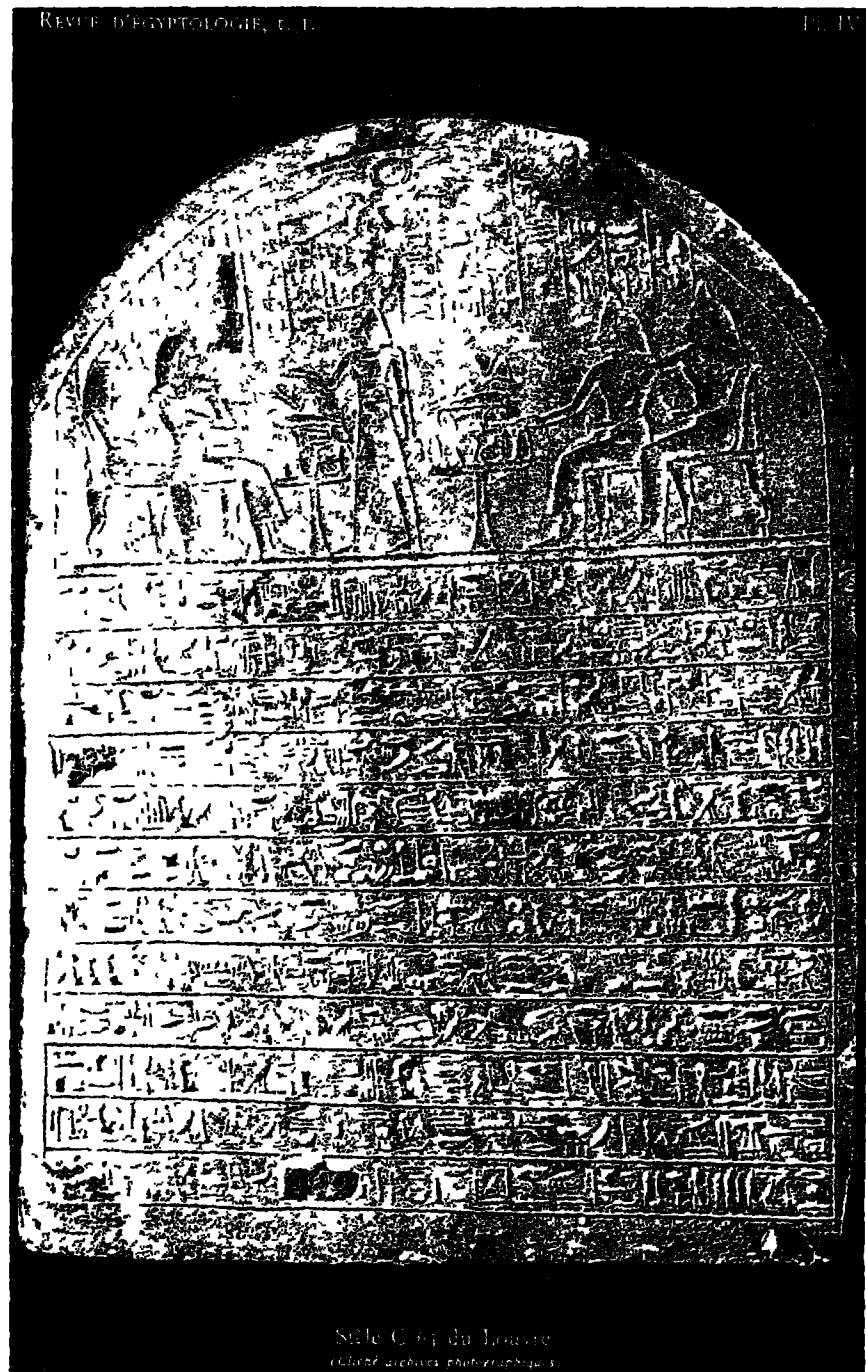
~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

Fig. 15-A.

~~CONFIDENTIAL~~

W F FRIEDMAN

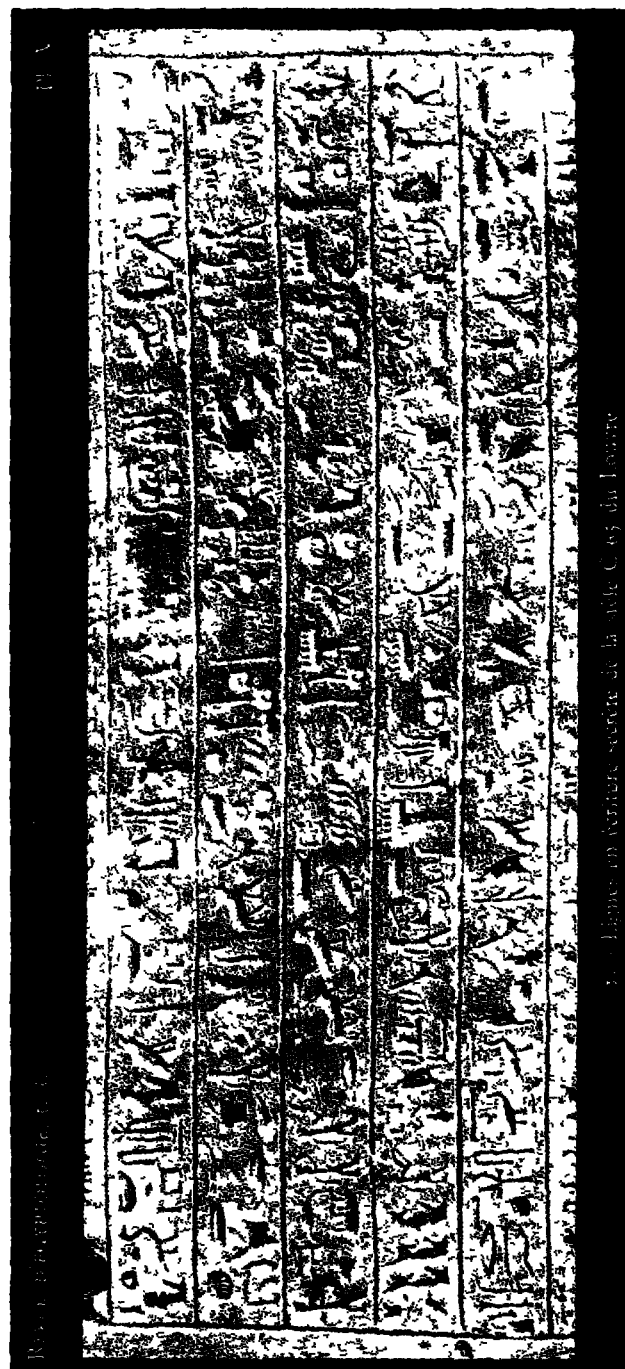
~~CONFIDENTIAL~~

Fig. 15-B.

~~CONFIDENTIAL~~

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

"The opening lines are filled with a list of the titles of Ptolemy V, and a series of epithets which proclaim the king's piety towards the gods, and his love for the Egyptians and his country. In the second section of the inscription the priests enumerate the benefits which he had conferred upon Egypt, and which may be thus summarized

- 1 Gifts of money and corn to the temples
- 2 Gifts of endowments to temples
- 3 Remission of taxes due to the Crown
- 4 Forgiveness of debts owed by the people to the Crown
- 7 Reduction of fees payable by candidates for the priesthood
- 8 Reduction of the dues payable by the temples to the Crown
- 13 Forgiveness of the debts owed by the priests to the Crown
- 14 Reduction of the tax on byssus (a kind of flax or cotton fibre)
- 15 Reduction of the tax on corn lands

Could it be that installment-plan buying was rampant in Ancient Egypt too, so that people didn't have enough left to pay their taxes?

Now, let's go back to those cryptographic hieroglyphics mentioned a moment ago. Here, in Fig 15-A for instance, is a picture of an inscription on a stela now in the Louvre, in Paris. Lines 6-10, inclusive, below the seated figures under the arch, contain secret writing in hieroglyphics, in Fig 15-B, these lines are seen enlarged. I won't attempt to explain the nature of the cryptography involved. It's pretty simple—something like the sort of cryptography involved in our own type of rebuses, and in our modern acronymic abbreviations, such as CARE, which stands for Cooperative (for) American Relief Everywhere, or NASA, for the National Aeronautics (and) Space Administration. Just to show you a bit of the cryptography that Drioton presents, without undertaking to explain what is involved, in Fig 15-C can be seen in sequence 34 hieroglyphic characters which are in lines 1, 2 and part of 3, of Fig 15-B (the 6th, 7th, and part of the 8th lines of Fig 15-A).

The following extracts, translated from a long article by Prof. Étienne Drioton in "Revue D'Égyptologie," Paris, 1933, will be of interest:

(P 1) "From the time of the Middle Empire onwards, Egypt had, alongside the official and normal system of writing, a tradition of cryptographic writing, the oldest known examples of which are to be found in the tombs of Beni-Hassan, and the most recent in the inscriptions of the temples of the Greco-Roman epoch

\* \* \* \* \*

(P 32) It is necessary to add to the enumeration of the cryptographic procedures the variation in the appearance of the cryptographic signs themselves. This variation, without however

~~CONFIDENTIAL~~

affecting their value, can (1) modify the appearance of the signs, (2) affect their position in various ways, and (3) combine these signs with others. Finally, to note a last peculiarity of these inscriptions which, because of their fine form, deserve to be considered the classics of the cryptography of this period, the scribe has several times successfully carried out in them what was doubtless considered to be the triumph of the genre the grouping of signs which offer a possible but fallacious meaning in clear, alongside a cryptographic meaning which is the only true one"

\* \* \* \* \*

And now for the most intriguing explanation offered by Drioton as to why cryptography was incorporated in these inscriptions. You know quite well why cryptography is employed in military, diplomatic, banking, and industrial affairs, you also know perhaps that it is used for other purposes, in love affairs, for example, and in illicit enterprises of all sorts, and you probably also know that it is often used for purposes of amusement and diversion, in tales of mystery, in the sorts of things published in newspapers and literary journals—they are called "crypts". But none of these explanations will do for the employment of cryptography in Egyptian hieroglyphics. Here's what Drioton thinks

(P 50) "There remains, therefore, the supposition that, far from seeking to prevent reading, the cryptography in certain passages of these inscriptions was intended to encourage their reading

The appeals which often introduce formulae of this type, and which are addressed to all visitors to the tombs, show in fact how much the Egyptians desired to have them read, but also, by the very fact of their existence, what an obstacle they encountered in the indifference, not to say satiety, produced by the repetition and the monotony of these formulae. To attempt to overcome this indifference by offering a text whose appearance would pique curiosity, based on the love, traditional in Egypt, for puzzles, to get people to decipher, with great difficulty, what was desired they should read, such is perhaps, in last analysis, the reason why the three monuments of the period of Amenophis III here considered present certain passages in cryptography

One must suppose, in this case, that the goal was not attained and that it was very quickly seen that the expedient produced, on the apathy of the visitors, an effect opposite to that intended. It removed even the slightest desire to read the inscriptions presented in this form. The new procedure was therefore, — the monuments seem to prove it —, abandoned as soon as it had been tried."

\* \* \* \* \*

~~CONFIDENTIAL~~

Before leaving the story of Champollion's mastery of Egyptian hieroglyphic writing, I think I should re-enact for you as best I can in words what he did when he felt he'd really reached the solution to the mystery. I'll preface it by recalling to you what Archimedes is alleged to have done when he solved a problem he'd been struggling with for some time. Archimedes was enjoying the pleasures of his bath and was just stepping out of the pool when the solution of the problem came to him like a flash. He was so overjoyed that he ran, naked, through the streets shouting "Eureka! I've found it, I've found it." Well, likewise, when young Champollion one day had concluded he'd solved the mystery of the Egyptian hieroglyphics, he set out on a quick mile-run to the building where his lawyer brother worked, stumbled into his brother's office, shouting "Eugene, I did it!", and flopped down to the floor in a trance where he is said to have remained immobile and completely out for five days. "Champollion died on 4 March 1832, leaving behind the manuscript of an *Egyptian Grammar* and of a *Hieroglyphic Dictionary* which, except for some errors of details inevitable in a gigantic work of decipherment and easily correctable, form the basis of the entire science of Egyptology"—Drioton, "*Decipherment of Egyptian Hieroglyphics*", *La Science Moderne*, August 1924, pp 423-432.

I shouldn't leave this brief story of the cryptanalytic phases of the solution of the Egyptian hieroglyphic writing without telling you that there remain plenty of other sorts of writings which some of you may want to try your hand at deciphering when you've learned some of the principles and procedures of the science of cryptology. A list of thus-far undeciphered writings was drawn up for me by Professor Alan C. Ross, of London University, in 1945, and had 19 of them. Since 1945 only two have been deciphered, Minoan Linear A and Linear B writing. The Easter Island writing is said to have very recently been solved, but I'm not sure of that. There are some, maybe just a very few, who think the hieroglyphic writing of the ancient Maya Indians of Central America may fall soon, but don't be too sanguine about that either.

Should any of you be persuaded to tackle any of the still undeciphered writings in the list drawn up by Professor Ross, be sure you have an authentic case of an undeciphered language before you. Figure 16 is one that was written on a parchment known as the Michigan Papyrus. It had baffled certain savants who had a knowledge of Egyptology and attempted to read it on the theory that it was some sort of variation—a much later modification—of Egyptian hieroglyphic writing. These old chaps gave it up as a bad job. Not too many years ago, it came to the attention of a young man who knew very little about Egyptian hieroglyphics. He saw it only as a

~~CONFIDENTIAL~~



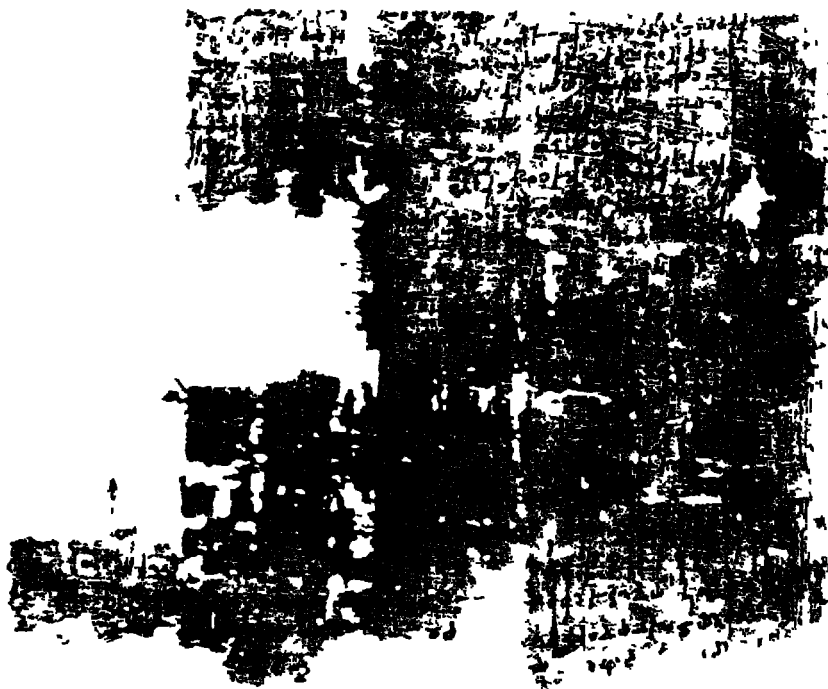


Fig. 16.

simple substitution cipher on some old language. He tackled the Michigan Papyrus on that basis and solved it. He found the language to be early Greek. And what was the purport of the writing? Well, it was a wonderful old Greek beautician's secret formula for further beautifying lovely Greek young women—maybe the bathing beauties of those days, among whom possibly were "Miss Greece of 500 B C" and "Miss Universe" of those days!

The next period of importance in this brief account of the history of cryptology is the one which deals with the codes and ciphers used by the contestants in our Civil War, the period 1861-65. It is significant and important because, for the first time in history, rapid and secure communications on a large scale became practicable in the conduct of organized warfare and world-wide diplomacy. They became practicable when cryptology and telegraphy were joined in happy, sometimes contentious, but long-lasting wedlock.

There is one person I should mention, however, before coming to the period of the Civil War in U S history. I refer here to Edgar Allan Poe, who in 1842 or thereabouts, kindled an interest in cryptog-

~~CONFIDENTIAL~~ HISTORY OF CRYPTOLOGY

raphy in newspapers and journals of the period, both at home and abroad For his day he was certainly the best informed person in this country on cryptologic matters outside of the regular employees of Government departments interested in the subject

In regard to Poe, one of our early columnists, there's an incident I'd like to tell you about in connection with a challenge he printed in one of his columns, in which he offered to solve any cipher submitted by his readers He placed some limitations on his challenge, which amounted to this—that the challenge messages should involve but a single alphabet In a later article Poe tells about the numerous challenge messages sent him and says "Out of perhaps 100 ciphers altogether received, there was only one which we did not immediately succeed in resolving This one we demonstrated to be an imposition—that is to say, we fully proved it a jargon of random characters, having no meaning whatever " I wish that cipher had been preserved for posterity, because it would be interesting to see what there was about it that warranted Poe to state that "we fully proved it a jargon of random characters " Maybe I'm not warranted in saying of this episode that Poe reminds me of a ditty sung by a character in a play put on by some undergraduates of one of the colleges of Cambridge University, in England At a certain point in the play, this character steps to the front of the stage and sings

"I am the Master of the College,  
What I don't know ain't knowledge "

Thus, Poe What he couldn't solve, he assumed wasn't a real cipher—a very easy out for any cryptologist up against something tough

If any of you are interested sufficiently to wish to learn something about Poe's contributions to cryptology, I refer you to a very fine article by Professor W K Wimsatt, Jr, entitled "What Poe Knew About Cryptography", Publications of the Modern Language Association of America, New York, Vol LVIII, No 3, September 1943, pp 754-79 In it you'll find references to what I have published on the same subject

This completes the third lecture in this series In the next one we shall come to that interesting period in cryptologic history in which codes and ciphers were used in this country in the War of the Rebellion, the War Between the States, the Civil War—you use your own pet designation for that terrible and costly struggle

~~CONFIDENTIAL~~