An Introduction to IBM Methods in Cryptanalysis

BY LAMBROS D. CALLIMAHOS

~~*Confidential*~~

A basic exposition of the principles of punched-card methods and their applications in cryptanalysis.

GENERAL

Electrical tabulating machines are widely used in commerce and industry to reduce the labor involved in processing or analyzing a large volume of data, or in performing a large number of manipulations or tests on a limited amount of data, or both. The fundamental idea inherent in tabulating equipment such as that developed by the International Business Machines Corporation is the recording of data in the form of holes punched in a card; the holes are used to establish timed electrical circuits which control the functions of the various machines through which the cards are fed. The cards are stacked in a hopper of an IBM machine and made to pass between a set of wire brushes and a brass roller; the presence of a hole in a column of the card permits the brush pertaining to that column to make contact with the roller, thus completing a circuit and operating an electromagnet. This closing of an electrical circuit at a definite time during the passage of a card through a machine and from a fixed position on the card is the basis upon which the various electrical tabulating machines function. The great flexibility of treatment afforded by the IBM system lies in the ability to arrange and rearrange the basic data in a form convenient for study, or to associate the basic data with other units of information on the same or different cards for the purpose of printing the data, comparing items, accumulating totals, etc.

THE IBM CARD

The cards used in the IBM system are 3¼" x 7⅜" and are made of specially prepared paper stock, strong in wearing quality and free from foreign particles which might act as conductors of electricity; moreover, the edges have been impregnated with a hardening composition which retards fraying of the edges under use, and thus makes for longer life of the card. Each card has 80 numbered columns in each of which may be punched a single item of alphabetical or numerical information

The opinions expressed in this article are those of the author(s) and do not represent the official opinion of NSA/CSS.

in either plain or coded[1] form. The columns contain 12 punching levels or positions; of these, 10 are indicated by the printed digits 0 to 9 in the horizontal lines on the card. The 11th and 12th punching positions, commonly referred to as the "x" and "y" punches, respectively, are at the top of the card and are not indicated by printed numerals. A hole in one of the levels from 0 to 9 results in the recording of that particular numerical datum, whereas a hole in an "x", "y", or 0 punch (called "zone punches") in conjunction with a numerical punch in the same column results in the coding of one of the 26 letters of the alphabet. The "x" and "y" punches by themselves have control functions in certain machine operations; furthermore, these punches are also used for the coding of special characters on the Tabulator such as a comma, an asterisk, or other symbols. In Fig. 1, below, is illustrated an IBM card,[2] slightly reduced, with alphabetical and numerical data punched in specific groups of columns designated as "fields[3]." Often it is convenient to use cards specially printed with

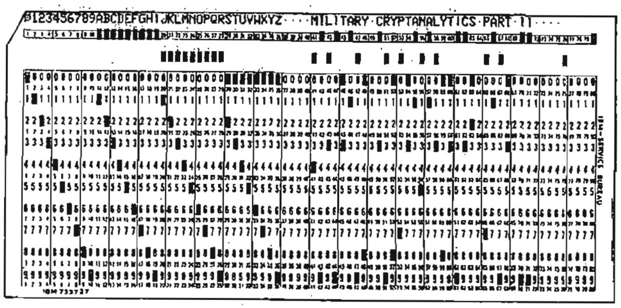

0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZ····MILITARY·CRYPTANALYTICS·PART·II····
IBM 733727

Fig. 1

vertical lines to indicate the fields; in addition, the designation of the field may be printed on the card. Each field defines a section of the card in which one particular type of information will always appear, and it is assigned a sufficient number of columns to include the largest number of alphabetical or numerical units which it will be called upon to accommodate.

FUNCTIONS OF THE PUNCHED CARD

The IBM system, employing a number of different machines taken collectively to constitute a working unit, makes possible multiple uses of a record in the form of a punched hole or a set of holes in the IBM card. A symbol punched in a card can be processed through many operations successively, since the functions of the various machines are rearranged for those operations through proper wiring of their control panels. Specifically, a single punched hole in a card may cause one or more of such operations as the following:

It may add the data represented by it to some other data;

It may subtract the data represented by it from some other data;

It may multiply the data represented by it by some other data;

It may divide the data represented by it into some other data;

It may cause the data represented to be listed (i. e., printed);

It may cause the data represented to be suppressed;

It may reproduce the data represented into a different field of the same card, or on a different card;

It may cause the data represented to be classified, or to be sorted;

It may cause the data represented to be selected;

It may cause the data represented to be printed on the IBM card;

It may cause the data represented to produce an automatic balance forward;

It may cause the data represented to be filed properly among other data;

It may cause a paper form to feed to a predetermined position or to be ejected automatically, or to space from one position to another; or

It may cause a total to be printed at the end of a group or class of data.

THE MACHINES IN THE IBM SYSTEM

The principal machines which constitute the IBM system are the following:

The *Card Punch,* with which the cards are punched with the information to be recorded;

[1] By "coding" in this connection is meant a conventional representation of data that may readily be adapted to IBM techniques. For example, the numbers 01 to 48 in a specified pair of columns could be used to represent the 48 States; or the letters "M", "P", "C", or "T" in a particular column might stand for "monoalphabetic", "polyalphabetic", "code", and "transposition", respectively.

[2] The printing at the top of the columns was performed simultaneously with the punching by a particular type of card punch equipped with this capability; in normal practice, punched cards do not contain such printing. The reader may find it instructive to cover the printing above cols. 41–71 and read the information with the help of the IBM coding shown in cols. 1–40.

[3] Note in this case the field comprised by cols. 1–40, and that comprised by cols. 41–75; also note the "x" punch in col. 76, and the "y" punch in col. 80.

The *Verifier*, which verifies the accuracy of the previously punched cards through an operation similar to key punching;

The *Sorter*, which sorts the cards in any groupings or classes desired (referred to as "major", "intermediate", and "minor" *sorts*);

The *Reproducer*, which reproduces extra copies of a deck of cards, with a rearrangement or selection of the punched data as desired;

The *Collator*, which performs various operations of merging, filing, matching, and selection of cards, and which may also be used to check the sequence of a sorted deck of cards;

The *Calculating Punch*, which performs operations of multiplication and division, and punches the result directly on the card from which it senses the data, or on a designated following card;

The *Interpreter*, which prints recorded data directly on the card from which it "reads" the information, as an aid in certain manual filing operations; and

The *Tabulator* (also called the *Printer*), which performs accounting operations of various classes of totals, and lists the data on continuous forms known as "IBM runs" or "IBM listings."

In addition to the foregoing standard machines there are special machines for specific purposes, e. g., a card-operated electric typewriter, machines for converting the punched holes of IBM cards into teleprinter tape punched with the coding of the Baudot code or vice versa (for use in certain semi-automatic procedures), etc. Moreover, each of the standard machines has available a number of extra devices which may be incorporated for specialized operations, according to the needs of the problems at hand.

GENERAL CRYPTOLOGIC APPLICATIONS OF THE IBM SYSTEM

Since electrical tabulating equipment is primarily designed as *data processing* machinery, we may use IBM machines in cryptanalysis to facilitate and expedite the examination of a large amount of traffic, in order to isolate homogeneous cryptosystems, or to prepare a group of homogeneous messages in many ways suitable for studying the patent or latent phenomena in the traffic and thus assist us in arriving at some conclusions regarding its cryptographic aspects, etc. All the manual work incidental to the solution of a cryptosystem could be performed by machine methods, but it must be emphasized that by no means can we assume that the era of clerical work is a thing of the past. Feasibility, practicability, and efficiency are the points which will determine what part of the work, or whether all of it, or none of it, should be done by machine techniques, and it is the IBM crypto-technician (i. e., an IBM expert with a broad background in cryptology) who is best qualified to advise in this respect. Although many phases of analysis are possible with machines, sometimes a judicious proportion of machine work plus manual clerical labor will permit the solution of a particular problem in the least possible time—certainly an important consideration in practical operational cryptanalysis.

Much of the preliminary groundwork that is necessary in the cryptanalysis of various types of cryptosystems may be done by machine methods. For example, we can find all repetitions within a message or among a set of messages and indicate by an asterisk those polygraphic repetitions which exceed a prescribed length; we can prepare message prints or work sheets arranged in proper groupings of characters (e. g., by period-lengths in the case of repeating-key ciphers, by digraphs in the case of digraphic systems, etc.); we can obtain frequency counts for single letters, digraphs, etc., of the over-all text, or columnar frequency counts of the text considering it to be written out on various trial widths; we can search for specified idiomorphs and print only those sequences corresponding to the specified patterns together with their location in the traffic, or we can search for isomorphic sequences and list their location; we can complete the plain-component sequence, both in monoalphabetic and polyalphabetic ciphers, and print the scores only of those generatrices which equal or exceed a certain predetermined threshold on the basis of two-category, arithmetic, or logarithmic weights; we can perform the ϕ test on distributions and print only those results which meet with pre-established minimum standards; we can match distributions on the basis of the χ or other tests; we can, once an additive has been recovered, remove the additive from superenciphered text and convert a message into monoalphabetic terms; and finally, having recovered all the keys to one message or a few messages in a particular cryptosystem, we can decrypt by machine methods the rest of the traffic in that same cryptosystem.

In the cryptanalysis of the more complex cryptosystems, there exist many advanced machine methods and techniques, some of which involve the applications of specialized equipments. A discussion of these methods will be left for possible future articles in the *Journal*.

IBM techniques are admirably suited to assist in traffic analysis studies. Traffic analysis is so closely affiliated with operational (as distinguished from academic) cryptanalysis that it is sometimes difficult to define where the one leaves off and the other begins—in present-day practice, these two fields of communication intelligence are complementary. In time of war when there is available for study a large volume of traffic emanating from many stations allocated into a plurality of radio nets, the data from intercept logs and other traffic analysis records may be rearranged in various major, intermediate, and minor sorts on elements such as the transmitting and receiving call signs, the file date and time, the group count, the external message numbers, indicator groups (if these be in the clear), priority or secrecy classifications, etc. The use of IBM runs for such studies greatly facilitates the grouping of stations which belong to the same net, selecting traffic that is cryptographically homogeneous, identifying and locating duplicate messages or isologs, finding communication "service" messages[4] which are so important to the cryptanalyst, identifying proforma messages or messages containing stereotyped reports, etc.

In the compilation of codes and ciphers, IBM methods furnish invaluable assistance. Through the use of machine techniques, code books are easily compiled several editions in advance of current needs; random alphabets for strip systems may be generated; and random keys for various cipher systems may be produced with facility and in volume. The use of IBM equipment eliminates the large expenditure of time and labor that is concomitant with the production of cryptosystems by manual methods.

APPLICATION IN A TYPICAL EXAMPLE IN CRYPTANALYSIS

Let us assume we have available for study the following five messages,[5] intercepted on the same day on a low-echelon ground net:

Message No. 1

```
RNZ  DE  LBF    5980KCS  180730Z

NRPWH FNDWU RMBNO KBFMJ WGHWM WEZLV UDOIN FJPGK DLASW
HHHZN INUFP EQVWB RSBMQ HKEWN UQXKH ZHBHD NVREE ESZBW
WRHEZ TTDTX
```

Message No. 2

```
UZK  DE  RNZ    5980KCS  180855Z

OWYWU ZNDQI AHIWB RRSJG IGHXN LCNFL THTNJ FJRZT WUZSF
OJWVH TQBAP TGNHJ KQADF GZBLL LXMQX EJHOK BPTRJ GDYIS
NRLKQ RLRRU RZEPB IVCMC ENHAX MIZQL KDRAS WTNGK EWNUZ
JXCJQ WIBGQ EZPDZ IQMPE HNHJK QRTKF ENBQD XZFVP HLQDT
YAVRE ZAPQF FOJJQ KWCMK HZHBH DNVR
```

Message No. 3

```
ZVH  DE  UZK    5980KCS  180920Z

MSHZH LASWH FGPEB ABJSX JKMLP MCIRE NWPJI RLMWY WBESC
SACJA TLFXY FGZUJ YMNJX CJYWK MCHAF BPZAH QNWJE WRAIQ
XJQVD LLRQU WJKKN KBGAE JUDLL WCMTG QDRA
```

Message No. 4

```
RNZ  DE  VGM    5980KCS  181000Z

KCEWN UNDRS JZYHF NSFLC BNHBE BKJWG KDTRH QVNVQ PQIMQ
IASJB LLLOT NWZKB YQRJJ QVDDI YTLRM ECWKS XJPGK ZZRWJ
IKEEE SALTJ ZDIPL RLMUC WIJZW DTPEH JJPWQ JGCCB EBRIQ
ZAFUH LWFGT ALJUT ZRCSR EJQXM WRIAG UGFYQ BJIGN DLLFC
H
```

Message No. 5

```
RNZ  DE  LBF    5980KCS  181145Z

DQEEW UKBFM JWGHW MNQSJ CBEBR DTBQV SPBIA ADQOC JLKQP
TNWZU LGIFO NLXHX XPBPE ZEHWX BAMRH YQZLS XMQXZ ULSPQ
XMGKP ZHSBD RDHHX RIQZN JMLLL IWODL KQ
```

The first thing to be determined is whether or not the messages are cryptographically homogeneous, i. e., in the same general system and specific keys. In the absence of indicators from which conclusions could be drawn, we must search for repetitions between messages, as well as for repetitions within individual messages; this is accomplished by preparing an *IBM index*, to be described below.

We will begin by punching each message on cards. The intercept data (and preamble information, if any), together with a sequential number (often called the "worksheet number") assigned to each message, are punched on a single card known as a "heading card."[6] The message text[7] is punched on "line cards" containing a predetermined maximum number of text characters per card; in this case, the line cards will contain 25 letters, except for possibly the last line card. (In addition to the message text, line cards will contain a reference line-letter or number, as well as the message number to which the

[4] A service message is a message between communications personnel pertaining to any phase of traffic handling, cryptographic operations, communication facilities, or circuit conditions.

[5] In actual practice, it would hardly be worth while to process five short messages by machine methods. It will be understood that this exposition applies to a larger volume of traffic.

[6] These heading cards are used for traffic analysis studies. It is customary to mark heading cards with an "x" punch in a specified column, in order to distinguish them from line cards in various machine processing operations.

[7] Indicators, if any, might be punched on the heading card; in any case, indicator groups would be deleted in the subsequent indexing procedure.

line cards pertain.[8]) Thus, in recording the data from the five messages, we will have a total of 5 heading cards and 31 line cards; these cards are then checked for accuracy of punching on the Verifier.

Next, by means of a process known as "offset gangpunching" involving the use of the Collator and the Reproducer, the 31 line cards are expanded into a deck of 736 cards, representing one card for each letter of the five messages. For instance, cols. 21–45 of the first line card of Message No. 1 contain the first 25 letters of the cipher text, *viz.*,

NRPWHFNDWURMBNOKBFMJWGHWM

while in the following "detail card" generated from the foregoing line card (including picking up the 26th letter of the cipher text from the second line card), we shall have[9]

RPWHFNDWURMBNOKBFMJWGHWMW

in cols. 21–45. The next detail card after that, picking up two cipher letters from the second line card, will contain the following letters in cols. 21–45:

PWHFNDWURMBNOKBFMJWGHWMWE

Thus for the 100 ciphertext letters of Message No. 1 we shall have generated 100 cards, each letter of the message appearing in col. 21 of a particular card. In addition to the letters of the cipher text, each card will contain further data as to the message number, and also the position in the message occupied by the letter appearing in col. 21 of the card.[10] There are further processing symbols incorporated into the cards, as an aid to subsequent machine treatment.

Using the Sorter, the 736 offset cards from the five messages are now put into alphabetical order according to tetragraphs. This operation takes four alphabetic sorts[11] on four consecutive columns going from right to left, as for example cols. 24–23–22–21. In other words, a "minor" sort is done on col. 24, "intermediate" sorts on cols. 23 and 22, and a "major" sort on col. 21. When this is finished, the sorted

[8] Information common to several line cards is punched automatically in the line cards from a prepunched master card inserted in the "duplicating rack" of the Card Punch.

[9] Note how each letter is offset one position to the left.

[10] The message number from the line cards is gangpunched in all the detail cards; the position number is punched in the cards by an automatic numbering device in the Reproducer.

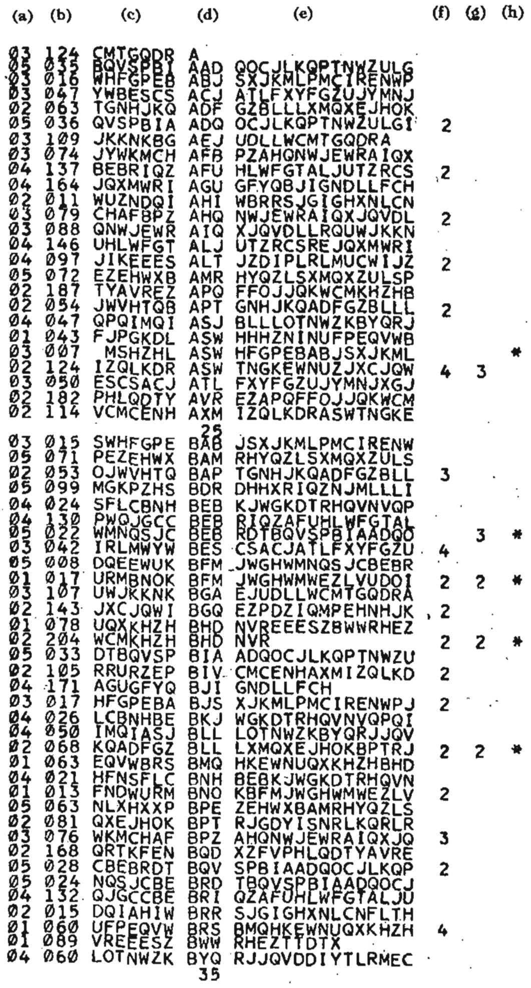
[11] In sorting alphabetical information, it is necessary to sort each column twice, once for the numerical punch and once for the zone punch; in numerical sorting a single sort per column suffices. Thus a tetragraphic sort on literal text requires 8 successive sorting operations.

deck is put into the hopper of the Tabulator, the control panel of which has been properly wired for this particular processing operation. The listing which results is known as an "IBM single-position index," or, more simply, an "IBM index"; the first page of this index is illustrated in Fig. 2. Every letter of the five messages appears as the first letter (called the "control letter") of the trigraphs in the column labeled "d"; the message number and the position of the control letter in the message are found in the columns labeled "a" and "b", respectively. The 7 letters which precede the control letter are listed in the column labeled "c", while the 15 letters which follow the control trigraph are listed in the column labeled "e". In order to facilitate examination of the IBM run, the control letter and the two following letters are here separated as an independent trigraph from the line of 25 letters; this, however, is an arbitrary convention in this particular case, since any spacing could be used as desired in the listing, the spacing not being in any way dependent upon the arrangement of the punched data on the card.[12] The digraphic and trigraphic frequencies ("intermediate totals" and "minor totals", respectively) are here printed at the right on the listing, in columns "f" and "g", respectively, whereas the uniliteral frequencies ("major totals") are recorded at the end of each grouping of A's, B's, etc., of the control letter.[13] For example, *it will be seen in the run that there are* 25 A's in the five messages; there are five cases of digraphic repetition in the "A" block with two occurrences each, as well as one case in which a digraph $(\overline{AS}_c)$ occurs four times; and there are three occurrences of the trigraph $\overline{ASW}_c$. An asterisk (*) in the column labeled "h" is here employed to indicate a tetragraphic or longer repetition, as may be seen in the case of the sequence $ASWH_c$.

At this stage the IBM single-position index would be sent to the

[12] It will be recalled that the cipher letters are punched in a solid block of 25 columns; the spacing desired is accomplished through proper wiring of the Tabulator control panel. It is of course also possible to rearrange the data punched on the cards in various ways in the listing; this again is dependent upon the wiring of the control panel.

[13] The kinds of totals, as well as the particular location in the printed listing of these totals, are governed by the wiring of the Tabulator control panel.



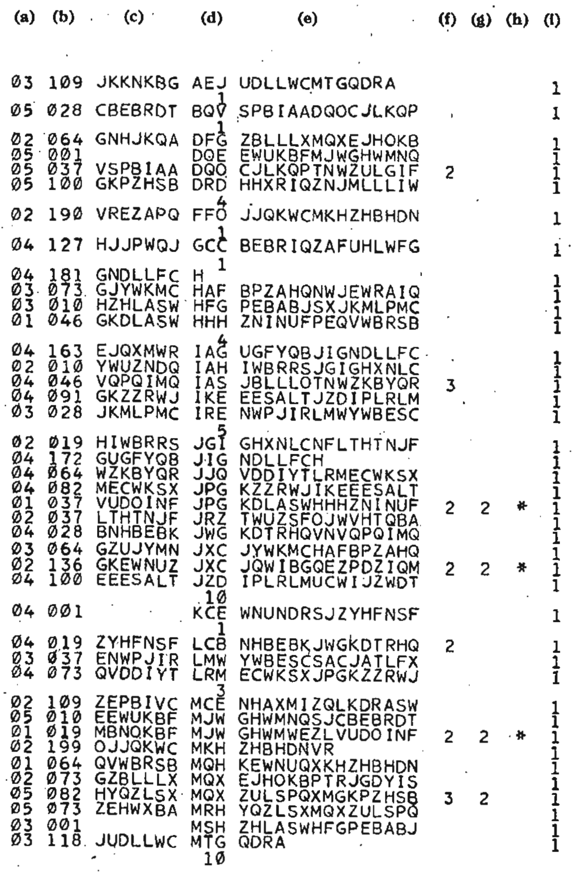
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
03	124	CMTGQDR	A				
05	035	BQVSPBI	AAD	QOCJLKQPTNWZULG			
03	016	WHFGPEB	ABJ	SXJKMLPMCIRENWP			
03	047	YWBESCS	ACJ	ATLFXYFGZUJYMNJ			
02	063	TGNHJKQ	ADF	GZBLLLXMQXEJHOK			
05	036	QVSPBIA	ADQ	OCJLKQPTNWZULGI	2		
03	109	JKKNKBG	AEJ	UDLLWCMTGQDRA			
03	074	JYWKMCH	AFB	PZAHQNWJEWRAIQX			
04	137	BEBRIQZ	AFU	HLWFGTALJUTZRCS	2		
04	164	JQXMWRI	AGU	GFYQBJIGNDLLFCH			
02	011	WUZNDQI	AHI	WBRRSJGIGHXNLCN			
03	079	CHAFBPZ	AHQ	NWJEWRAIQXJQVDL	2		
03	088	QNWJEWR	AIQ	XJQVDLLRQUWJKKN			
04	146	UHLWFGT	ALJ	UTZRCSREJQXMWRI			
04	097	JIKEEES	ALT	JZDIPLRLMUCWIJZ	2		
05	072	EZEHWXB	AMR	HYQZLSXMQXZULSP			
02	187	TYAVREZ	APQ	FFOJJQKWCMKHZHB			
02	054	JWVHTQB	APT	GNHJKQADFGZBLLL	2		
04	047	QPQIMQI	ASJ	BLLLOTNWZKBYQRJ			
01	043	FJPGKDL	ASW	HHHZNINUFPEQVWB			
03	007	MSHZHL	ASW	HFGPEBABJSXJKML			*
02	124	IZQLKDR	ASW	TNGKEWNUZJXCJQW	4	3	
03	050	ESCSACJ	ATL	FXYFGZUJYMNJXGJ			
02	182	PHLQDTY	AVR	EZAPQFFOJJQKWCM			
02	114	VCMCENH	AXM	IZQLKDRASWTNGKE			
			25				
03	015	SWHFGPE	BAB	JSXJKMLPMCIRENW			
05	071	PEZEHWX	BAM	RHYQZLSXMQXZULS			
02	053	OJWVHTQ	BAP	TGNHJKQADFGZBLL	3		
05	099	MGKPZHS	BDR	DHHXRIQZNJMLLLI			
04	024	SFLCBNH	BEB	KJWGKDTRHQVNVQP			
04	130	PWQJGCC	BEB	RIQZAFUHLWFGTAL			
05	022	WMNQSJC	BEB	RDTBQVSPBIAADQO		3	*
03	042	IRLMWYW	BES	CSACJATLFXYFGZU	4		
05	008	DQEEWUK	BFM	JWGHWMNQSJCBEBR			
01	017	URMBNOK	BFM	JWGHWMWEZLVUDOI	2	2	*
03	107	UWJKKNK	BGA	EJUDLLWCMTGQDRA			
02	143	JXCJQWI	BGQ	EZPDZIQMPEHNHJK	2		
01	078	UQXKHZH	BHD	NVREEEESZBWWRHEZ			
02	204	WCMKHZH	BHD	NVR	2	2	*
05	033	DTBQVSP	BIA	ADQOCJLKQPTNWZU			
02	105	RRURZEP	BIV	CMCENHAXMIZQLKD	2		
04	171	AGUGFYQ	BJI	GNDLLFCH			
03	017	HFGPEBA	BJS	XJKMLPMCIRENWPJ	2		
04	026	LCBNHBE	BKJ	WGKDTRHQVNVQPQI			
04	050	IMQIASJ	BLL	LOTNWZKBYQRJJQV			
02	068	KQADFGZ	BLL	LXMQXEJHOKBPTRJ	2	2	*
01	063	EQVWBRS	BMQ	HKEWNUQXKHZHBHD			
04	021	HFNSFLC	BNH	BEBKJWGKDTRHQVN			
01	013	FNDWURM	BNO	KBFMJWGHWMWEZLV	2		
05	063	NLXHXXP	BPE	ZEHWXBAMRHYQZLS			
02	081	QXEJHOK	BPT	RJGDYISNRLKQRLR			
03	076	WKMCHAF	BPZ	AHQNWJEWRAIQXJQ	3		
02	168	QRTKFEN	BQD	XZFVPHLQDTYAVRE			
05	028	CBEBRDT	BQV	SPBIAADQOCJLKQP	2		
05	024	NQSJCBE	BRD	TBQVSPBIAADQOCJ			
04	132	QJGCCBE	BRI	QZAFUHLWFGTALJU			
02	015	DQIAHIW	BRR	SJGIGHXNLCNFLTH			
01	060	UFPEQVW	BRS	BMQHKEWNUQXKHZH	4		
01	089	VREEESZ	BWW	RHEZTTDTX			
04	060	LOTNWZK	BYQ	RJJQVDDIYTLRMEC			
			35				

Fig. 2

cryptanalyst, who would study it to see what he could see.[14] He would note that there are many polygraphic repetitions between messages, many more than would be expected by mere chance; this is proof that the traffic is homogeneous, supporting the probability of homogeneity that was implied from traffic analysis information. He would also note that there is a pentagraphic repetition in Message No. 2, at an interval of 99 (obtained from the run by subtracting the numerical position of the first occurrence from that of the second occurrence), suggesting factors of 3, 9, or 11. On examining the intervals of the repetitions *between* messages, he would find that 9 is a factor common to all the repetitions, proving that the messages are in flush depth[15] and indicating that the cryptosystem is probably a polyalphabetic cipher of 9 alphabets.

[14] It may be pointed out that the cryptanalyst at this time is really not sure of what he expects to see in this first index. The index was ordered as a means of examining the *over-all* immediate phenomena associated with the uniliteral, digraphic, and trigraphic frequencies, as well as finding all polygraphic repetitions present in the cipher text; i. e., the cryptanalyst hopes by this first index to find evidence of *nonrandom* characteristics in the cipher text or in its over-all frequencies. In addition to these obvious elements (which constituted the main reason for which the index was ordered), any other characteristics which appear to be *nonrandom* would be searched for by the cryptanalyst. For example, in a particular case he might note that polygraphic repetitions occur in the main only between messages originating from the same *transmitting station*, or between messages transmitted within certain periods of time, or between messages having identical or nearly identical elements in the preamble such as serial numbers or other groups, etc.

[15] If the messages had been offset on a keying cycle of 9, the intervals of repetitions between messages could have been factored to 9 if a constant equal to the offset were first subtracted from the interval; this could have been noted from multiple repetitions between two messages.



```
(a) (b)    (c)    (d)      (e)           (f) (g) (h) (i)

Ø3 1Ø9 JKKNKBG AEJ UDLLWCMTGQDRA                 1
                 1
Ø5 Ø28 CBEBRDT BQV SPBIAADQOCJLKQP               1
                 1
Ø2 Ø64 GNHJKQA DFG ZBLLLXMQXEJHOKB               1
Ø5 ØØ1         DQE EWUKBFMJWGHWMNQ               1
Ø5 Ø37 VSPBIAA DQO CJLKQPTNWZULGIF   2           1
Ø5 1ØØ GKPZHSB DRD HHXRIQZNJMLLLIW               1
                 4
Ø2 19Ø VREZAPQ FFO JJQKWCMKHZHBHDN               1
                 1
Ø4 127 HJJPWQJ GCC BEBRIQZAFUHLWFG               1
                 1
Ø4 181 GNDLLFC H                                 1
Ø3 Ø73 GJYWKMC HAF BPZAHQNWJEWRAIQ               1
Ø3 Ø1Ø HZHLASW HFG PEBABJSXJKMLPMC               1
Ø1 Ø46 GKDLASW HHH ZNINUFPEQVWBRSB               1
                 4
Ø4 163 EJQXMWR IAG UGFYQBJIGNDLLFC               1
Ø2 Ø1Ø YWUZNDQ IAH IWBRRSJGIGHXNLC               1
Ø4 Ø46 VQPQIMQ IAS JBLLLOTNWZKBYQR   3           1
Ø4 Ø91 GKZZRWJ IKE EESALTJZDIPLRLM               1
Ø3 Ø28 JKMLPMC IRE NWPJIRLMWYWBESC               1
                 5
Ø2 Ø19 HIWBRRS JGI GHXNLCNFLTHTNJF               1
Ø4 172 GUGFYQB JIG NDLLFCH                       1
Ø4 Ø64 WZKBYQR JJQ VDDIYTLRMECWKSX               1
Ø4 Ø82 MECWKSX JPG KZZRWJIKEEESALT               1
Ø1 Ø37 VUDOINF JPG KDLASWHHHZNINUF   2   2   *   1
Ø2 Ø37 LTHTNJF JRZ TWUZSFOJWVHTQBA               1
Ø4 Ø28 BNHBEBK JWG KDTRHQVNVQPQIMQ               1
Ø3 Ø64 GZUJYMN JXC JYWKMCHAFBPZAHQ               1
Ø2 136 GKEWNUZ JXC JQWIBGQEZPDZIQM   2   2   *   1
Ø4 1ØØ EEESALT JZD IPLRLMUCWIJZWDT               1
                1Ø
Ø4 ØØ1         KCE WNUNDRSJZYHFNSF               1
                 1
Ø4 Ø19 ZYHFNSF LCB NHBEBKJWGKDTRHQ   2           1
Ø3 Ø37 ENWPJIR LMW YWBESCSACJATLFX               1
Ø4 Ø73 QVDDIYT LRM ECWKSXJPGKZZRWJ               1
                 3
Ø2 1Ø9 ZEPBIVC MCE NHAXMIZQLKDRASW               1
Ø5 Ø1Ø EEWUKBF MJW GHWMNQSJCBEBRDT               1
Ø1 Ø19 MBNQKBF MJW GHWMWEZLVUDOINF   2   2   *   1
Ø2 199 OJJQKWC MKH ZHBHDNVR                      1
Ø1 Ø64 QVWBRSB MQH KEWNUQXKHZHBHDN               1
Ø2 Ø73 GZBLLLX MQX EJHOKBPTRJGDYIS               1
Ø5 Ø82 HYQZLSX MQX ZULSPQXMGKPZHSB   3   2       1
Ø5 Ø73 ZEHWXBA MRH YQZLSXMQXZULSPQ               1
Ø3 ØØ1         MSH ZHLASWHFGPEBABJ               1
Ø3 118 JUDLLWC MTG QDRA                          1
                1Ø
```

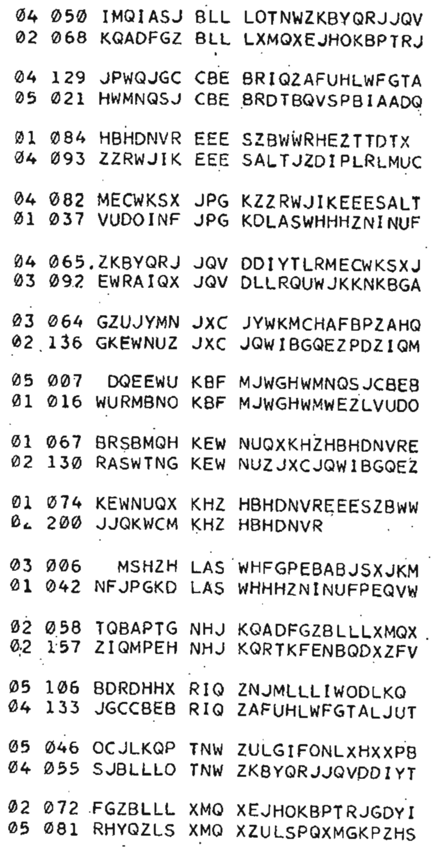
Fig. 3

With the periodicity established as 9, work sheets of the messages are prepared according to this period; this step may be performed by IBM if the number of messages warrants, without the necessity for further manual card punching. The first over-all IBM index has served its purpose, that of aiding the cryptanalyst in locating repetitions and diagnosing the cryptosystem; so now a new index is prepared, to assist in solution of the plain text. This second run (the first page of which is illustrated in Fig. 3) is an index *by alphabet* prepared by first sorting the cards back into their original sequence in each message, collating these cards cyclically into 9 blocks (each containing all the cards belonging to one particular alphabet), and performing a tetragraphic sort on the blocks;[16] the deck is then listed on the Tabulator, yielding what amounts to an elaborate triliteral frequency distribution showing many prefix and suffix letters of the control trigraph, in addition to the frequencies by alphabet of the single letters, digraphs, and trigraphs. This index shows, for example, that there are 10 J's in the first alphabet, and indicates two sets of polygraphs that begin with the letter J in Alphabet 1. With this second run the cryptanalyst can proceed to locate all the *causal* repetitions in the five messages—information which, when coupled with the statistical information given in the index, will considerably simplify solution of the cryptosystem.

FURTHER REMARKS

In case the cryptanalyst is interested only in *totals* of various classes (uniliteral, digraphic, etc.), condensed listings showing only such data may be prepared by a process known as "tabulating", at twice the speed of ordinary listing. If, on the other hand, the condensed data are to be arranged in, let us say, descending order of frequency, a procedure known as "summary punching" is employed. In this case a Reproducer is connected electrically to the Tabulator; totals accumulated (and perhaps printed) by the Tabulator are punched by the Reproducer on "summary cards", together with the information associated with the totals. This technique is often employed when certain information is desired without the volume of data that normally goes along with the usual IBM index.

[16] The appropriate alphabet number is gangpunched in each block; this alphabet number is listed in the column labeled "i" of the run.



```
Ø4 Ø5Ø IMQIASJ BLL LOTNWZKBYQRJJQV
Ø2 Ø68 KQADFGZ BLL LXMQXEJHOKBPTRJ

Ø4 129 JPWQJGC CBE BRIQZAFUHLWFGTA
Ø5 Ø21 HWMNQSJ CBE BRDTBQVSPBIAADQ

Ø1 Ø84 HBHDNVR EEE SZBWWRHEZTTDTX
Ø4 Ø93 ZZRWJIK EEE SALTJZDIPLRLMUC

Ø4 Ø82 MECWKSX JPG KZZRWJIKEEESALT
Ø1 Ø37 VUDOINF JPG KDLASWHHHZNINUF

Ø4 Ø65.ZKBYQRJ JQV DDIYTLRMECWKSXJ
Ø3 Ø92 EWRAIQX JQV DLLRQUWJKKNKBGA

Ø3 Ø64 GZUJYMN JXC JYWKMCHAFBPZAHQ
Ø2 136 GKEWNUZ JXC JQWIBGQEZPDZIQM

Ø5 ØØ7  DQEEWU KBF MJWGHWMNQSJCBEB
Ø1 Ø16 WURMBNO KBF MJWGHWMWEZLVUDO

Ø1 Ø67 BRSBMQH KEW NUQXKHZHBHDNVRE
Ø2 13Ø RASWTNG KEW NUZJXCJQWIBGQEZ

Ø1 Ø74 KEWNUQX KHZ HBHDNVREEESZBWW
Ø2 2ØØ JJQKWCM KHZ HBHDNVR

Ø3 ØØ6   MSHZH LAS WHFGPEBABJSXJKM
Ø1 Ø42 NFJPGKD LAS WHHHZNINUFPEQVW

Ø2 Ø58 TQBAPTG NHJ KQADFGZBLLLXMQX
Ø2 157 ZIQMPEH NHJ KQRTKFENBQDXZFV

Ø5 1Ø6 BDRDHHX RIQ ZNJMLLLIWODLKQ
Ø4 133 JGCCBEB RIQ ZAFUHLWFGTALJUT

Ø5 Ø46 OCJLKQP TNW ZULGIFONLXHXXPB
Ø4 Ø55 SJBLLLO TNW ZKBYQRJJQVDDIYT

Ø2 Ø72 FGZBLLL XMQ XEJHOKBPTRJGDYI
Ø5 Ø81 RHYQZLS XMQ XZULSPQXMGKPZHS
```

Fig. 4

If there are a large number of messages to be indexed, and if all the cryptanalyst desires is to find all repetitions of a specified length or longer, it is possible to obtain a condensed run giving only the necessary information, suppressing the large volume of unwanted data. For example, let us say that in the first listing (Fig. 2) all we wish to find are tetragraphic or longer repetitions; such a listing is illustrated in Fig. 4, wherein are shown all repetitions of the specified lengths for the set of five messages. Note that four of the asterisked repetitions shown in Fig. 2 have been suppressed in Fig. 4; only the long repetitions which *begin in the column of the control letter* are included in the listing of Fig. 4—this of course cuts down on redundant information, and simplifies even more the examination of the run.[17]

The examples shown of IBM indexes have been *single-position* indexes. Where the cryptographic unit consists of two or more characters, indexes are modified accordingly. For example, if a cryptosystem involved a four-letter code, then a *four-position* index would be made of the traffic, showing a control column of four-letter code groups with one or more preceding and following four-letter groups.

In some cases, it might be advisable to prepare indexes with the major sort on message number; in other cases, an index might be prepared of a number of messages grouped together by indicator relationships, or grouped by date or period of transmission, etc. When a comparison is made between all messages in a large volume of traffic to determine possible homogeneity or relationships, the index resulting from this operation is known colloquially as a "brute force." Such an index makes exorbitant demands on available machine time with standard machines; but on the other hand, *in extremis* it might be the only guarantee of solution in a complex cryptosystem.

In conclusion, it is hoped that the reader has gained an appreciation of the nature of the aid that can be given by IBM techniques in cryptanalysis; but at the same time he should also realize that there may be a delicate point of transition where manual methods leave off and *IBM methods begin*, or vice versa. This is especially true when the volume of messages or other data to be examined is small.

[17] The way in which this run was prepared may be of interest. The tetragraphically sorted deck is passed at tabulating speed in the Tabulator, without printing. Whenever a tetragraphic coincidence is noted by the comparison brushes in the Tabulator, the machine senses the letter immediately to the left of the control letter of the two tetragraphs; if these letters are identical, it shows that the repetition extends in that direction, and will therefore be picked up again at an offset; but if these letters are different, it shows that the repetition *begins in the control column*, and the Tabulator will print the lines belonging to this tetragraphic repetition. Another way in which this could be accomplished is by using the Collator to remove cards with the proper coincidences from the sorted deck, and then listing only these cards on the Tabulator; this latter procedure might be used in situations involving a large volume of cards, to take advantage of the high speed of operation of the Collator.