An Advanced Problem in Cryptography and Its Solution

By

1st Lieutenant J. O. Mauborgne
24th U. S. Infantry

Preliminary Note

THE following problem and solution is printed for the information of the student officers of the Army Signal School and officers on the mailing list of the school. Its solution has been unsuccessfully attempted by several officers interested in work of this character and their lack of success can now be pointed out as due to failure to apply scientific principles and sufficient imagination to the problem.

With the following analysis and solution at hand, any Playfair cipher of reasonable length should be deciphered in time.

It was not necessary for the author to determine whether or not the cipher was a Playfair, so that a few remarks on this preliminary point may be of interest to complete the analysis.

If an unknown cipher be examined and the vowels are found to be approximately 40% of the whole, the cipher may safely be considered as a transposition cipher and should be handled as such. If the vowels form a small per cent of the whole and the less common letters, as X, J, Z, Q, occur frequently, the cipher may be safely classed as a substitution cipher, of which the Playfair is one variety. On dividing the unknown substitution into groups of two letters each, examine the groups and see if any group consists of a repetition of the same letter, as S S. If so the cipher is not a Playfair. If no repetition groups are found, the cipher may or may not be a Playfair. Examine it now for group frequency, i. e. see if groups such as UG, DX, etc., occur again and again. If so the cipher may safely be considered a Playfair. Of course, groups in any substitution cipher will recur, but in the Playfair, they recur with great frequency and the groups which do not recur may be in the minority. P. H.

THE following cryptogram, without any additional data, was handed to the writer almost two years ago, by Lieut. Karl Truesdell, Signal Corps, who had received it from its author, Captain M. Muirhead, R.F.A., the writer of several excellent articles on Military Cryptography which have appeared in the Technical Conference proceedings of the Army Signal School within the last few years:

"Playfair's Cipher"

UGNES	GHFEF	DXCVI	KCNMY	PTUGH	DEFPH	QFDEN	VAYRE	MXBUD	PNDPS
OPLNF	BNYUG	ZUUBQ	XPTRL	DQIDG	QWUFI	NSSGN	VUZEF	UGDEN	LRISP
SPSZP	TUGHD	EFPHQ	NSVUV	IYFHM	XNHQS	NRRNZ	RIMNH	UGNDB	IUYUG
ZUVUN	PSZLO	IDUVH	SPSUZ	VEVXP	HUYUG	CXDRP	MHDEF	DPSNP	MPTRL
DCLDO	YSPXN	FDNDP	SVXLI	VSOXK	MFILX	NCFMP	SENLN	BYQNF	IXDBI
GULIE	PDHHE	NRFIF	BMRTV	PHUYU	GNEHU	VUDHP	MPTRL	DCIBO	UXPDL
DCZSX	DUGID	XEPEI	ZGRDM	DPLVV	SOXKN	FILXN	CFMPS	ENLNV	YQYPT
RXLDE	FDXPY	IKFIP	QDXZU	VPVCH	SPOPN	DKIDD	QCNDN	YUPDV	WSZUG
DELNR	ISPSO	HDOYR	DUYXV	QXPEU	YXVFN	SPGYL	IOPWU	KMFIB	EHUVU
DHSVO	YHIHT	OYFIL	XNENC	INFDT	ULIVO	LYHFE	FDXCV	IKCNZ	BWZPE
CYRGS	VRNVO	EFRNO	UNLDC	QXEHH	RDXPH	QOHFL	CUGNR	UZXVS	ZRZZB
EPYSP	SNERD	DPBZX	VXEXD	DNXEP	EVNVT	QSNIV	PYPLD	OQZUI	KNRXD
ENCID	LDSSG	UGNES	GHFEF	DXCVI	KCNAN	CYPAI	KUGDR	ZUXVP	LBFEK
IDXRF	YUZXV	ZXLIU	HOXCE	XHDPA	EDYAN	IWGXC	NOPZS	RIPDU	VVMFW
UGDRN	DWZFI	GQZYL	IRNIN	SEPSU	GDXVH	GUDBC	MPSFP	XOEAV	RZYPH
XOCID	XFIFB	XYDXA	YREMX	BFHNU	YUVVM	UGDLN	AANIW	GXCNP	NHSPO
DLFWI	DIKVQ	XPRN-	"						

The students of this school were practically challenged by Captain Muirhead to discover the meaning of the cryptogram, but a solution, so far as the writer knows, has never been forthcoming from any of those to whom the cryptogram was given, until the writer, during the first leisure time he has had in several years, while enroute to the Philippine Islands, attacked the problem by the method given, and found the correct solution.

The line of reasoning employed illustrates the application of the principles laid down in the excellent articles by the author of the cryptogram, and, as it may act as a guide to student officers at the Signal School and elsewhere, it is here reproduced in full.

The actual time spent in arriving at a solution was almost twelve hours, though the work was actually spread out over the leisure time of three days.

Captain Muirhead considerably shortened the time and labor of deciphering the cryptogram by the fact that he labelled it "Playfair's Cipher." This is the standard British field cipher which is not entirely unknown in the United States, and, as a knowledge of its principles of operation was one step in the solution of the above message, and probably can be used under similar conditions by other officers of the army in the future, it is here reproduced so that the reader can see to what extent it may be used in the solution of a cryptogram enciphered by it, when the key word is unknown, as it is in this case.

Playfair's Cipher

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

A key word is used in which no letter is repeated, as "GRANT." A large square of twenty-five smaller squares is constructed as shown, the letters of the key word being written in order from left to right starting in the upper left hand corner of the large square, and followed by the remaining letters of the alphabet in regular sequence, until the square is completely filled up. It is customary to consider I and J as one letter in this cipher so they usually are written together in the same square.

The text of the message to be sent is then divided up into groups of two letters each, and equivalents are found for each pair.

Every pair of letters in the square must be:

(1) In the same vertical line. Thus in the above example each letter is represented in cipher by that which stands next below it, and the bottom letter by the top one of the same column; for instance, EY is represented by LN.

(2) In the same horizontal line. Each letter in this case is represented by that which stands next on its right, and the letter on the extreme right by that on the extreme left of the same horizontal line with it; for instance DF is represented by EB.

(3) At opposite corners of a rectangle. Each letter of the pair is represented by the letter in the other corner of the rectangle in the same horizontal line with it; for instance BU is represented by FO.

If, on dividing the letters of the text into pairs, it is found that a pair consists of the same letter repeated, a dummy letter, as X, Y, or Z, should be introduced to separate the similar letters.

If the message to be sent were "The enemy moves at dawn," after being divided into pairs it would be enciphered thus,

TH EX EN EM YM OV ES AT DA WN
GM DY LE FL ZL VG LY NG KD YR

The message is then broken up into groups of five letters for transmission.

To decipher such a cryptogram, (knowing the key word), the receiver divides it into pairs, and from his table finds the equivalent of these pairs, taking the letter immediately above each, when they are in the same vertical line; those immediately on the left, when in the same horizontal line; and those at opposite angles of the rectangle when this is formed.

With this knowledge in view, the writer's first step in the solution of the cryptogram was to divide the whole message into pairs of letters so that he could arrive at the frequency of the pairs, leaving sufficient space between lines, so that it was possible to interline the deciphered letters, thus:

UG	NE	SG	HF	EF	DX	CV	IK	CN	MY	PT	UG	HD	EF	PH
QF	DE	NV	AY	RE	MX	BU	DP	ND	PS	OP	LN	FB	NY	UG
ZU	UB	QX	PT	RL	DQ	ID	GQ	WU	FI	NS	SG	NV	UZ	EF
UG	DE	NL	RI	SP	SP	SZ	PT	UG	HD	etc.				

Before proceeding to examine the frequency of the pairs, however, the writer noticed the following recurring combinations of letters in the message:

UGNESGHFEFDXCVIKCN appears twice.
PTUGHDEFPHQ appears twice.
VSOXKMFILXNCFMPSENLNVYQ appears twice.
EFDX appears four times.
AYREMXB appears twice.
HFEFDXCVIKCN appears three times.

All of these recurring groups were underscored for future consideration.

An indexed blankbook was then used to note down under proper alphabetical heading, all of the pairs of letters in the cryptogram, indexing them in the order in which they occurred in the text, and when they were found to recur, making a check mark for each time they appeared. It was found that this is a less tiresome and far quicker method of making this count than to take one pair, for example, "UG,"

and go through the entire message to see how often it occurred, and then, to do the same thing for every other pair in turn.

Not only was the frequency of each pair noted but also the frequency of the reverse of each pair. The frequency of some of the more frequently recurring pairs is here given. The writer, of course, found the frequency of all the pairs but it is not necessary to tabulate all of them in this paper.

UG	17	GU	2
EF	9	FE	0
DX	9	XD	4
FI	9	IF	0
PS	8	SP	5
IK	7	KI	0
CN	6	NC	3
PT	6	TP	0
PH	6	HP	0
UY	6	YU	1
XV	6	VX	2
LI	6	IL	0
ID	6	DI	0
DP	5	PD	0
NE	5	EN	3
RN	5	NR	4
SP	5	PS	8
ZU	5	UZ	4

To quote from Captain Muirhead's articles on Cryptography to be found in the proceedings of the Technical Conferences of the Army Signal School for the years 1911-12 and 1912-13, "the most frequent pairs occurring in English are:

TH	149	ON	81	ND	61	HA	52	TO	45
HE	119	RE	75	AT	58	EN	49	TI	45
AN	84	IN	63	OF	55	NT	49	ST	44
ER	81	ED	61	OR	53	EA	45	IT	40

The most frequent triplets occurring in English are:

THE	67	HAT	14	FOR	13	NDE	11	NCE	10
AND	21	EDT	13	ION	13	HAS	10	OFT	10
THA	18	ENT	13	TIO	12	MEN	10	STH	10

Also, the ordinary table of frequency of single letters in English is as follows:

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

The letters in brackets recur an equal number of times."

As UG in the message occurred more often than any other pair it is undoubtedly the equivalent of TH, and, from the nature of Playfair's Cipher, GU will, conversely, be HT; so these pairs are written above the corresponding pairs of the cryptogram wherever they occur.

The alphabetical index of the pairs was again referred to, and it was found that the frequency of the *initial letters of the pairs* was as follows:

A	6	F	21	L	20	Q	10	V	23
B	8	G	8	M	5	R	17	W	4
C	16	H	18	N	28	S	19	X	21
D	40	I	21	O	14	T	2	Y	3
E	17	K	3	P	37	U	33	Z	13

J did not appear at all.

Now, while it may be seen, from the construction of Playfair's Cipher, that it is possible to have the initial letters of a pair represent several different letters of the text, still these initial letters will combine most frequently in such a manner as to represent one certain letter, and we may argue that the frequency of the initial letters of the pairs is practically that of the single letters of the language. With that in mind, we find that initial D occurs most frequently, and, consequently, it probably represents E in the majority of combinations. Moreover we note that DX occurs 9 times to 5 of DP, 4 of DL, etc.

Now the most frequently recurring pair in English that begins with E is ER, so we substitute throughout the body of the cryptogram ER for DX, and RE for XD.

Well down toward the end of the message we find the following group:

UG	DX	VH	GU	which by substitution equals
TH	ER	—	HT	

It does not take much exercise of the imagination to cause one to decide that VH is probably the equivalent of IG; so the group means THE RIGHT, and we begin to feel that we are on the right trail.

Referring again to the frequency of the initial letters of the pairs, we decide that, as D, occurring 40 times, probably represents E in the majority of combinations, and U occurring 33 times represents T, so P with a frequency of 37, will, in the majority of combinations, represent O, and initial V occurring 23 times will, perhaps, be found to represent I, since its frequency, by comparison, is of about that order in the text. Moreover, it has been found in the combination VH to represent IG. Since initial G has been found to most frequently represent H, it might have been expected that VH would represent IG. Reasoning in a similar manner, if V is most frequently I, and U is most frequently T, there is a strong probability that VU will be found to equal IT, and UV to equal TI. Also, by the same process of reasoning, we may infer that VP will represent IO, and PV, OI. Similarly, TP ought to represent UO, and PT, OU. So we substitute all these various combinations in the text.

In the first line we find PTUG, which by substitution must represent OUTH, perhaps part of the word SOUTH. This combination also occurs in the third line of the message.

The constant recurrence of UGN in the message caused the writer to count up the triplets in the message, and it was found that UGN occurred most frequently; so the chances are that UGN is equal to THE, and we see that in many cases in the text, N must equal E.

An effort to construct the alphabetical square by which the message was enciphered, from the data already deduced, was next made. The writer de-

cided that the square, to be correctly arranged, must satisfy the conditions already found, viz.:

(1) UG equals TH
(2) DX equals ER
(3) VH equals IG
(4) VU equals IT
(5) VP equals IO
(6) TP equals UO
(7) Initial N often equals E

The first step was to draw a diagram like a checkerboard on a piece of cardboard, and in each of the 25 squares to write the proper letter of the alphabet which would appear in that square if no code word were used, J being omitted, as it did not appear in the cryptogram.

The letters were placed in the lower right hand corner of each square.

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

Twenty-five little squares of cardboard, slightly smaller than the squares of the fixed diagram, were cut out and marked each with a letter of the alphabet, corresponding to the fixed letters of the diagram, the object being to place these letters on the diagram, and move them around until the arrangement of all the letters found, viz., U, G, T, H, D, X, E, R, V, I, T, O, P, and N, satisfied the conditions found to exist in the part of the message already deciphered.

With these conditions satisfied, it was believed that other letters of the alphabet would fall naturally into their proper places, and that one pair after another would be found which would further add to the deciphered part of the message, and that all the letters of the alphabet would ultimately be so arranged that the key word would disclose itself.

A few points about the mechanism of the Playfair Cipher occurred to the writer at this time. The letters of the alphabet which move least from their normal positions in the alphabetical square, owing to the rearrangement necessary when a key word is used, are those in the lowest part of the square. With a great many key words the last five letters of the alphabet, V, W, X, Y, and Z, never move at all. On the other hand, was not this very pertinent fact, and the great assistance it would give to one engaged in solving a cryptogram, clearly understood by so clever a cryptographer as Captain Muirhead, and would he not, on that account, deliberately try to break up that lowest line by shifting some of the letters from it to the key word? If so, what letters would he be most likely to shift? Undoubtedly V, W, and Y could be utilized most easily, so X and Z were the only letters that seemed to have a fair chance of being left in the last line, and of these, only Z might be expected to remain in its usual position.

Further it may be stated that two vowels will usually be found in the key word.

The first move on the diagram was made by supposing that the last line had remained untouched, and that T and U retained their normal positions. Supposing that only two letters had been taken from between G and T for use in the key word, then G and H, which normally lie beside each other in the same line, would each move two spaces to the right,

and the first condition, that UG equals TH, would be satisfied.

From the statement DX equals ER it appeared that R had probably moved to the right until above X, so that to satisfy the conditions found in the message, E must have been above D, in other words in the first line, and therefore part of the key word, and both D and E must have been in the same vertical line with R and X, in order that the condition DX equals ER might be satisfied. The first two conditions might be fulfilled by the following arrangement:

		E		
		D	G	H
		R	T	U
V	W	X	Y	Z

But the fact that requirements 3 and 4, viz., that VH equal IG and VU equal IT would not be satisfied by this arrangement showed that it was not the correct one.

Moving U to the last line (supposing one letter to have moved from the bottom line to the key word) would not be correct, even if we moved H to a corresponding position above it on the third line, for the requirement that UG equal TH would no longer be satisfied, so it was concluded that two of the letters of the bottom line must have gone to the key word, permitting both T and U to take positions in the last line, and G and H must have been moved to positions

in the third line immediately above T and U. Supposing also that Z had not moved, and that X had probably moved one space to the right, since Y was a more likely candidate for the first line than V or W, the E, D, R, X combination was moved one space to the right, and the arrangement took this form:

			E	
			D	
G	H			
			R	
T	U		X	Z

Conditions 1 and 2 are satisfied, and we try the effect of placing first V and then W in the space left in the lowest line to see if conditions 3, 4 and 5 are satisfied in either case. With V in that space, the equation that VH equals IG is found unsatisfied, showing that the position of V is incorrect, and that it must, therefore, appear in the key word, and that W is probably the letter remaining in the last line. It may be argued that perhaps Y remained, and that the X column should not have been moved to the right. Later trials proved that the writer's hypothesis was the correct one.

Now we are certain that S has to follow R in line 4 of the square, and fairly certain that W follows U in the fifth line, so we put them in the spaces mentioned, and decide to find the proper place for both V and Y if it is possible to do so at the present state of the examination. Now conditions 3 and 4 are:

(3) VH equals IG and (4) VU equals IT.

Both of these equations will be satisfied if we

put VI in the first line above GH. We now have the following:

V	I		E	
			D	F
G	H			
			R	S
T	U	W	X	Z

F must precede G and follow D since E is to be found in the first line, or it is in the key word and so are Z and S, and the vertical X line would necessarily move one space to the right. But it is very improbable that Z, S, and F are in the key word, and a trial of the new pairs DS and FR in the cipher will be found to still further decipher the cryptogram, thus showing the correctness of the writer's assumption, as will be demonstrated almost immediately.

Requirements 5 and 6, namely, that VP equals IO, and TP equals UO will be satisfied if we write OP under GH, and we then see that Q must follow P; so our square now has this arrangement:

V	I		E	
			D	F
G	H			
O	P	Q	R	S
T	U	W	X	Z

To revert to the question of the correctness of the writer's assumption that F, S, and Z are now properly placed, referring to the cryptogram we find in the lowest third of it the group DSSGUGN. Now we know that UGN equals THE. DS, if F and S are correctly placed, is the equivalent of FR and we get by substitution FR—— THE. If we write M in the square above S, SG will represent OM, and substituting these letters in the group above makes sense, for we read FROM THE. So our assumption that F and S are properly placed seems to be correct.

Since M is followed in the alphabetical square by O, N must appear in the key word. There are only two spaces left between H and M, and, as I is in the key word and J is omitted, K and L must be the two missing letters in the third line, so our square now is arranged thus:

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

That leaves unplaced, the letters A, B, C, N, and Y, of which we are sure N and Y are in the key word, since we find them missing from their places in the alphabet.

The first twelve letters of the cryptogram are

UGNESGHFEFDX

Let us substitute as many letters as possible. Remembering that UGN equals THE, we get

U G N E S G H F E F D X
T h e — o m m — n d e r

We see that the missing letters in the group probably are C and A, and that the whole group should read THE COMMANDER. The pair NE, therefore, equals EC, and HF equals MA, so our square now contains three more letters:

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

Only Y and B remain, and, as Y is to be in the key word, it must follow N, and B will have to follow A. So the completed square is arranged thus:

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

Now we can substitute in the cryptogram and see if we have reasoned correctly.

TH EC OM MA ND ER IN CH IE FG OU TH LA ND HA
UG NE SG HF EF DX CV IK CN MY PT UG HD EF PH

SB E. EN IY LX LZ AW AR EF OR SO ME DA VF TH
QF DE NV AY RE MX BU DP ND PS OP LN FB NY UG

XT WA RW OU LD BR EA K—
ZU UB QX PT RL DQ ID G— etc.

After running along smoothly for a few words it begins to appear that we have made some mistake in the arrangement of our key square, and we discover that the letters Z and Y should change places. So the arrangement of the square will be:

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

Starting anew with the final adjustment of the square, and barring some accidental errors in the cryptogram, due, perhaps, to mistakes in copying the original, we arrive at the following translation:

"THE COMMANDER-IN-CHIEF SOUTHLAND HAS BEEN FULLY AWARE FOR SOME DAYS THAT WAR WOULD BREAK OUT ANY MOMENT, AND THE EMPEROR OF SOUTHLAND HAS CONTINUALLY IMPRESSED ON HIM THE FACT THAT IT IS OF GREAT IMPORTANCE THAT THE WELSHLAND ARMY SHOULD BE DESTROYED BEFORE THE NORTHLAND REINFORCEMENTS CAN REACH THEIR ALLIES, AND ALSO, THAT THE CAPITAL SHOULD BE CAPTURED BEFORE THE ARRIVAL OF LARGE NORTHLAND REINFORCEMENTS WOULD RENDER SUCH AN OPERATION IMPOSSIBLE. A BRIEF EXTRACT OF THE EMPEROR'S LAST LETTER WRITTEN FROM THE SOUTHLAND CAPITAL LAST AUGUST AND RECEIVED BY THE COMMANDER-IN-CHIEF AT BRINWOL ON SECOND SEPTEMBER WILL PERHAPS MAKE THE STATE OF AFFAIRS MORE CLEAR. AFTER REFERRING TO PRE-

VIOUS DISPATCHES RECEIVED FROM THE COM MANDER-IN-CHIEF, IN WHICH THE LATTER HAD CLEARLY STATED THE PARTICULAR DIFFICULTIES OF OPERATING BY THE LEFT BANK OF THE SEVERN, OR THE RIGHT BANK, OR ASTRIDE OF THAT RIVER, AND AFTER FULLY ADMITTING THE DIFFICULTIES IMPOSED BY EACH COURSE——"

The quotation is incomplete. There is only one doubtful point in the translation. Near the middle of the message we find the following group, S V O Y H I H T O Y, which is translated O N S T A U G U S T. The writer has translated this LAST AUGUST on the supposition that S V should have been H D. However, it is quite possible that a numeral stood between O N and S T in the original message, and was not enciphered. So there is a chance that the mesage should have read "ON 21st AUGUST," or some such date, which is really immaterial to us.

To the proposition "*Cryptograms of sufficient length, enciphered by Playfair's cipher, are decipherable in a comparatively short time, and present only moderate difficulty to their solution,*" we can now safely add "Q. E. D."