Annex 16. Notes on the Breaking of GEAM Using the 'Winds – Set-Up' Message

David Sissons

Greater East Asia Ministry (GEAM) was introduced on 21 July 1943. The *Report* on Japanese Diplomatic Cyphers describes it as follows:

> It was a transposition cypher with a bigram and tetragram code, the bigrams being consonant-vowel or vowel-consonant and the tetragrams made up of a double consonant and a bigram. All letters of the alphabet were used and Y was regarded as both consonant and vowel. The code was patterned after the manner of LA. Messages were transposed in blocks of ten by ten without blanks. There were 26 indicators each providing a column and a row order, the row order containing only nine figures as the bottom row of the cage was composed of dummies, namely the first letter of the indicator repeated ten times. The 26 keys were designated by the letters A through Z and each key was indicated by its letter and the following letter of the alphabet in the form ABABA. The indicator was located at the head of the cypher text.

In the race to break GEAM, Trendall had his solution completed by August 12th, beating his rivals from both London and Washington. From the account given in the report, and from the box-copies of Trendall's daily cables to London for that period (which miraculously survived among the office files of the Captain I(x) at Victoria Barracks), it is possible to piece together how he did it.

As the cyphertext of only one of the many thousands of messages intercepted by 52 Special Wireless Section has found its way to the National Archives of Australia (NAA), and that was encyphered in BA not GEAM, I have for the purposes of illustration selected a typical Japanese Foreign Ministry message[1] and encoded it using the copy of GEAM's code and encyphering instructions for GEAM Mark 1 that are held by the US National Archives.[2]

---

1 Circular 2353 of 11 November 1941, the famous 'Winds – Set-up' message, reproduced in D. Kahn, 'Pearl Harbor and the Inadequacy of Cryptanalysis', *Cryptologia*, October 1991, at pp 288–92.

2 US National Archives, Record Group 457, Box1328, File 190/37/34/3, 'US/UK Technical Exchanges and Information on Solution of JBB' (19 pp).

Specimen Message—Encoded

Kanchō fugō atsukai.
AC YK HU LY WA UF

Kokusai jōkyō no hippaku no kekka itsu saiaku no jitai ni
TU YU YH YC GO HI FU QA CU GO SU CA BI FU YU BA CU GO MI WY GI

tachiitaru ka mo hakararezaru tokoro kakaru baai waga hō to aite koku no
FA FI BI FA KU CA JO HA CA KA KE MA KU UQ CA CA KU PA TA ZZLO FO TA FE TU GO

tsūshin wa tadachi ni teishi serarubeki o motte waga hō no gaikō kankei kiken ni hinsuru
IR ID UR XXOH AW DI DE KA KU PE CI VVQI ZZLO GO RRAD AC ZI CI EC GI IH DU KU

baai ni wa waga kaigai hōsō no kakuchi muke nihongo news no chūkan oyobi saigo
PA TA GI UR UR LA YA VI HY DY GO ZA FI JU CE GI OH LO YQ DU GO YJ AC YE YU LO

ni oite tenki yohō to shite (1) Nichibei kankei no baai ni wa "higashi no kaze ame"
VVEL EF KI RO HY ZZBI CCVY WI UW AC ZI GO PA TA GI UR IN HI LA DI GO CA ME BA JE

(2) Nichiso kankei no baai ni wa "kita no kaze kumori" (3) Nichiei
UN CCWE WI DO AC ZI GO PA TA GI UR IN RI FA GO CA ME CU JO KI UN CCWY WI XA

kankei no baai (tai shinchū mare Netherlands East Indies kōryaku o fukumu) "nishi no
AC ZI GO PA TA NI WY ID YJ JA OT AK IB CY YX BO IZ JU NU IN GI DI GO UN BO

kaze hare" o 2 do zutsu kurikaeshi hōsō seshimeru koto to seru o motte
CA ME HA KE UN BO VU NO MU FU CU KI CA BE DI HY DY WWKU KU CO FO DE KU VVQI

migi ni yori angō shorui to tekitō shobun aritashi. Nao migi wa gen ni gokuhi
IW VVIK KI AB LY SO KU BI FO FE CI FY SO UP BA KI FA DI UK UX IW UR EL GI RREN

atsukai to seraretashi.
WA YA FO WWIK FA DI UK

In this example the Japanese cypher clerk would have selected the appropriate encyphering key (in this case FGFGE) and proceeded to enter the encoded text into a series of transposition blocks of 10 cells x 10 cells dimensions, row by row, in the row-order prescribed by the key FGFGE (which is: 1 2 3 4 5 9 8 7 6)

On completion, the first block would look like this:

6	W	Y	G	I	F	A	F	I	B	I
7	F	A	K	U	C	A	J	O	H	A
8	C	A	K	A	K	E	M	A	K	U
9	U	Q	C	A	C	A	K	U	P	A
5	Y	U	B	A	C	U	G	O	M	I
4	G	O	S	U	C	A	B	I	F	U
3	G	O	H	I	F	U	Q	A	C	U
2	U	F	T	U	Y	U	Y	H	Y	C
1	A	C	Y	K	H	U	L	Y	W	A
	F	F	F	F	F	F	F	F	F	F

He would then write out the message for transmission, column by column, in the column order prescribed by the key FGFGE (which is: 10 2 6 9 7 4 3 1 5 8).

Thus it would appear on his message pad, and on the message of the Australian telegraphist receiving it, as follows:

[Originator, Addressee, Number, Date, each encoded]									FGFGF
IOAUO	IAHYF	YAAQU	OOFCF	FJMKG	BQYLF	AAEAU	AUUUF	BHKPM	FCYWF
GKKCB	SHTYF	FCKCC	CFYHF	IAUAI	UUCAF	IUAAA	UIUKF	WFCUY	GGUAF
IURYO	EWROF	RCAAV	EREAF	ZDUHL	PAIFF	CHIIZ	UHOOF	CKUDG	CDITF
AGTYQ	KXTZF	AIGVZ	KOGLF	IURYO	IIDAF	DIAAI	AXUZF	REPLV	DUFTF
CIIEI	FEQUF	YONAO	OOIOF	AGDJB	EYYJF	WAAAZ	LCOIF	ZUGUC	KYDCF
WPHMH	VYOZF	UTLBZ	EALFF	IRONC	IUUEF	IAIEY	VJHAF	VGICR	LGGGF
AJYUI	OAAOF	CITON	ARICF	PYCNW	JFTDF	ODBUY	UIAIF	TJYIX	KGGAF
ZWAIC	MIGWF	GIIJW	CRPWF	AAXNA	IOICF	IYKZC	ENOEF	ANOBU	CUZCF
UKUYU	EUENF	OIBOY	UOAIF	KIKFK	BMKUF	EVOIW	AOAOF	VKBSK	DFUBF
FILFD	KVMDF	DVSCW	CNHGF	VIIOU	IUNOF	OWYEY	IUEIF	CQAFH	CBCGF
RAAXA	FKIOX	PFUWF	XFFWN	KXIFE	YDUDF	URWXB	FIEIX	KFLAI	KIFXR
WXAFU	GFXUF								

Trendall's technique of solution was the application of a sequence of processes, the first of which is is described in the report as follows

> A frequency count of a few messages showed that consonants and vowels were used in almost equal numbers. As we already had an example of a vowel-consonant, consonant-vowel code (LA), the theory was straightway suggested that GEAM was such a code transposed. The regular occurrence of the dummy letter at intervals of ten gave the probable length of the lines as ten. Moreover, as the dummies appeared at shorter intervals at the end of a message (when the final transposition form was incomplete) and these final dummies always began after a multiple of 100 letters, it was obvious that the size of the block was ten by ten.

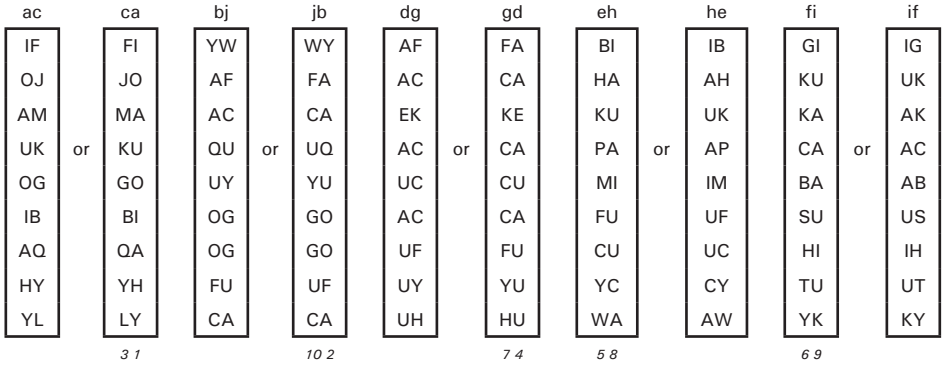
Applying this technique to our specimen message, an examination of the incoming message for periodicity reveals the Indicator (FGFGE) followed by the cryptogram in which the letter F appears at intervals of ten up to the 500th letter and thereafter at intervals of six. The latter suggests that the encoded message has been encyphered by transposing it in blocks of ten cells x ten cells until the concluding block (in which the columns are only six cells, instead of ten cells, high). If so, Block 1 consists of ten columns, each ten cells high, which for convenience we may identify by the letters a to j, thus:

a	b	c	d	e	f	g	h	i	j
I	Y	F	A	B	G	F	I	I	W
O	A	J	A	H	K	C	A	U	F
A	A	M	E	K	K	K	U	A	C
U	Q	K	A	P	C	C	A	A	U
O	U	G	U	M	B	C	I	A	Y
I	O	B	A	F	S	C	U	U	G
A	O	Q	U	C	H	F	U	I	G
H	F	Y	U	Y	T	Y	C	U	U
Y	C	L	U	W	Y	H	A	K	A
F	F	F	F	F	F	F	F	F	F

a	b	c	d	e	f	g	h	i	j
I	Y	F	A	B	G	F	I	I	W
O	A	J	A	H	K	C	A	U	F
A	A	M	E	K	K	K	U	A	C
U	Q	K	A	P	C	C	A	A	U
O	U	G	U	M	B	C	I	A	Y
I	O	B	A	F	S	C	U	U	G
A	O	Q	U	C	H	F	U	I	G
H	F	Y	U	Y	T	Y	C	U	U
Y	C	L	U	W	Y	H	A	K	A
F	F	F	F	F	F	F	F	F	F

The next task is to arrange these columns in their correct sequence. If in the GEAM code the bigrams were restricted to vowel followed by consonant and consonant followed by vowel it may be possible to fit the columns into their correct pairs on that basis. This Trendall proceeded to do. Let us apply this technique to our example, omitting of course, the terminal F dummies. Take Column *a* and put it beside each of the other columns in turn to see whether in any case it produces only bigrams of this restricted pattern. The result is that it will not pair with either Column *b* (that would produce adjoining vowels OA AU OU IO AO and adjoining vowels HF), Column *d* (adjoining vowels IA OA AA UA OU IA AU), Column *f* (adjoining consonants HT), Column *h* (adjoining vowels II OA AU UA IU AU, and adjoining consonants HC),Column *i* (adjoining vowels II OU AA UA OA IU AI), or Column *j* (adjoining vowels UU). It pairs, however, with Columns *c*, *e*, or *g*.

Continue this process with each column in turn. Column *b* pairs with Column *j*. Column *c* pairs with Column *a*. Column *d* pairs with Column *g*. Column *e* pairs with Column *h*. Column *f* pairs with Column *i*. Thus we have all five pairs:



From here, the report continues:

> It merely remained to put these groups together on repeats. This part of the job did actually present a few stumbling-blocks, as the presence of a row-order was not at first suspected and we were unfortunate enough to have constructed all the repeats backwards. The code-breakers were therefore somewhat baffled when these texts were presented to them—although it was later found that the main difficulty was the fact that the early messages were encoded from an English text. The first guide to the real solution was given by the incomplete blocks in which an incomplete line of five letters appeared at the end of the line instead of at the beginning. Thus we saw that we had our keys backwards; this was remedied and then long repeats were found going from one line to another, the second line not necessarily being the next in order. From then on the breaking of the code and row order was a relatively simple matter and the complete code and cypher system were known within a fortnight.

The next task is to discover which five of these ten possible bigram columns are the correct ones and their correct sequence. (The answer is indicated by the italicized figures that we have inserted beneath the bigram columns above). How on earth Trendall managed to do this beats me.

Where a cryptogram is merely the plaintext transposed, the sequence of the columns can be be established by anagramming, taking advantage of the characteristics and idiosyncracies of the mother tongue in which it is sent. A good example of this method is the solution of the

following English language cryptogram of a military report:[3]

3 L. D. Callimahos & W. F. Friedman, *Military Cryptanalysis,* part 2, vol. 2 (Aegean Park Press reprint, 1985), p.p. 418–20.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
N	R	E	O	U	N	M	P	L	E	T	C	A	O	N	Y	E
E	T	T	H	U	E	E	J	H	R	S	E	U	T	N	D	R
T	S	D	O	F	R	D	A	O	U	D	H	R	S	C	Y	E
E	S	R	E	I	O	S	O	E	V	D	N	A	S	G	R	S
F	T	H	E	R	T	H	E	P	A	S	T	E	O	E	R	T
L	F	X	A	R	C	A	G	M	A	O	L	E	S	M	N	H
T	M	S	D	R	F	R	E	W	I	E	H	N	Y	E	E	I
D	D	W	U	R	I	T	W	A	N	O	F	O	S	T	O	N
S	C	H	O	S	E	C	N	W	A	A	L	T	T	T	O	R

We note the J in column 8. In English, J is always followed only by a vowel, usually U. In this row columns 5 and 13 have a U. Let us provisionally postulate the latter and pair columns 8 and 13. In English the bigram JU must be followed by a consonant, usually N or S. These letters are present in columns 15 and 11. We now juxtapose columns 15 and 11 in turn against our 8–13 pair. This provides the alternatives

8	13	15
P	A	N
J	U	N
A	R	C
O	A	G
E	E	E
G	E	M
E	N	E
W	O	T
N	T	T

3	13	11
P	A	T
J	U	S
A	R	D
O	A	D
E	E	S
G	E	O
E	N	E
W	O	O
N	T	A

The English-language trigrams formed by columns 8–13–11 look more like plaintext trigrams than do those formed by columns 8–13–15. From here, the anagramming progresses rapidly, by expanding the trigrams into the words that have begun to manifest themselves, such as PATROL, JUST, ROAD, etc. The complete plaintext together transposition key emerges as below.

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

It was just in order to prevent anagramming of this nature that in the GEAM sytem the plaintext is *encoded* before transpositon. How is anagramming to be conducted when all one knows about the code is that it consists of vowel-consonant and consonant-vowel bigrams and tetragrams each representing a *kana* syllable or common word or phrase?

Trendall must have gone about identifying the code groups of particular syllables by the frequency of their appearance. Let us attempt to apply this method to our specimen message. Here the report's list of the most commonly appearing syllables in the Japanese language is of assistance: *no*, *ni*, *wa*, *o*, *shi*, *mo*, *to*. These six, presumably, are listed in their order of frequency in a large collection of diplomatic messages. In a message of 833 syllables that I selected from the *Pearl Harbor Hearings*,[4] the number of times each of these appeared was as follows: *no* 37, *ni* 15, *wa* 20, *o* 28, *shi* 12, *mo* 8, *to* 16.

The report says: 'It merely remained to put these groups together on repeats'. Which are 'these groups'? What is here meant by a 'repeat'? Does it mean the reappearance of a commonly occurring sequence of bigrams (e.g. VVSY AS ** LY VVDY = 'Reference my telegram No.**')? If so, it would not appear to be a technique applicable at this stage; for we do not yet know a single code group, or in which row the message starts. I'm afraid I've no alternative but to seek your [i.e., Kenneth McKay] assistance.

GEAM was introduced on July 21st. By a rare stroke of luck a complete file of your *outward* signals to London from August 5th (until the following April) has survived. Perhaps these cast some light on how it was done?

On August 5th and succeeding days you sent London the column order for about 5 keys each day.

4 United States Congress (79th Congress), Joint Committee on the Investigation of the Pearl Harbor Attack, *Pearl Harbor Attack; Hearings*, part 37, p. 988, Ambassador, Washington to Consul-General Honolulu N. 384 of 27 November 1941.

The following are some of your signals. I have inserted in square brackets the actual values taken from the code charts in Appendix L.

August 10th

Following bigram sequences have been established as frequent in more than one key: (a) BO [o] DI [shi] often preceded by KI [ri] or KU [ru]; (b) CU [ku] MU [zu] FU [tsu]; (c) FA [ta] KE [re] often preceded by CI [ki] and followed by UR [wa] BI [i]; (d) FO [to] CO [ko] KU [ru] (two of this group ususally occur in very close proximity; (e) FU [tsu] MI [ji]; (f) JO [mo] KU [ru] often preceded by GO [no] or FO [to] and followed by DU [su]; (g) KU [ru] DU [su] often preceded by JO [mo] RY [yu] and/or folllowed by FO [to] or CU [ku]; (h) KU (ru] GA [na] often preceded by LA [ga] and/or followed by AS [dai]; (i) TI [1] AS [dai]; (j) BO [o] KU [ru]; (k) KU [ru] MA [za] BE [e].

My comment is that, except for the fact that all of these are back-to-front, many of them make sense; for example, in (b) tsuzuku is a common word in such messages, meaning 'to continue'; in (k) ezaru is the negative potential inflexion meaning 'cannot'.

August 11th

1. GEAM cypher is based on bigram table similar in general arrangement to LA with which many groups are identical. Individual sentences are read backwards but present evidence suggests that successive sentences, as separated by full stops UF, should be taken forwwards.

2. Following groups are different from LA: SA zero, TI 1, VU 2, WE 3, XO 4, AR 5, IS 6, UT 7, EV 8, OW 9, UR wa, BO o, CO ko, DO so, FO to, GO no, LO go, RY yō, YE oyobi, CA kan, VI gai, ZASS kiden [your telegram], SYVV ōden [my telegram], DYVV ni kanshi, FUQQ aritashi.

3. LA Spelling Table is used apparently with OG and QU.

My comments are that these values are correct except for 'end spelling', which should be YY not QU, and for the tetragrams, which are back-to-front and should read SSZA, VVSY, VVDY, and QQFU.

August 12th

1. Further research shows that instruction to read backwards by sentences is incorrect and following should be substituted. Keep message in originl blocks consisting of 9 rows of 5 bigrams each. On keys so far supplied, each of these lines should be read backwards, but order of reading horizontal rows inside block varies according to key, which

> accordingly consists of two sets of figures, one for vertical columns and one for horizontal rows. Rows so far established are: ABABA 1 to 9 in serial order; XYXYX and FGFGF 6 7 8 9 5 4 3 2 1; OPOPO and QRQRQ 4 3 2 1 5 6 7 8 9. Blocks are read successively in natural order downwards.
>
> 2. In order to avoid nuisance of reading backwards we shall in future transpose column order in pairs. Thus ASASA column order will be 5 2 8 6 4 9 1 10 7 3 instead of present 7 3 1 10 4 9 8 6 5 2. Same bigrams will then read naturally forwards, but tetragrams will appear with double consonants first, thus VVFY instead of FYVV and we shall in future quote them in this form.
>
> 3. LA bigrams in series BA to PA, BI to PI, BU to PU, BY to PY, UB to UP, EB to EP, and OB to OP seem to remain unchanged. . .

In each of the days that followed you sent London large packets of bigrams and tetragrams that you had established.

From the above it would appear that: (i) You had the whole game completely sewn up by August 12th; (ii) You had worked out how to establish column orders by August 5th without knowing any code groups; (iii) Without knowing row orders you were, by August 10th, working out indicative frequent bigram strings merely on the basis of individual rows of 5-bigrams length and from this you were able confidently to establish the correct values of some bigrams and tetragrams by that date; (iv) Armed with this information you were then able to establish row orders. Have I got this right?