The ADFGVX Cipher

Wilson Poulter & Justin Kulp, *Lakehead University*

The ADFGVX cipher is a private-key encryption method that uses a Polybius square to encrypt a plaintext message once, it then uses a keyword to transpose letters of the singly encrypted text, adding additional difficulty for cryptanalysis.

History

On the frontline of trench warfare during the First World War (1914-1918), communication was an integral part of daily activities. In order to carry out an effective offensive, messages containing battle plans had to be transmitted along the kilometres of trench networks to commanding officers. The obvious method for accomplishing this task at the time was through the use of radio telegraphy. By these means, operators could send messages using Morse code to other operators located far away. Although this was the only practical way of sending messages at the time, the use of radio had the drawback that one could not control who the message was sent to. Thus, enemy operators could intercept virtually every message sent, providing them with a tactical advantage. In order to protect the content of messages even after they had been intercepted, cryptographers from all countries began to develop ciphers that would protect their messages even if they were intercepted [1].

The ADFGX cipher originally came into use on March 5, 1918, and later evolved into the more complicated ADFGVX cipher to include numbers. The method was invented by the German Colonel Fritz Nebel and was chosen by a conference of German cipher specialists to be used during the war. It was designed to optimize radio operator success as well as cryptanalytic security [2].

At the time, the French army had a dedicated cryptanalysis group known as the *Bureau du Chiffre*, or Cipher Bureau. By April 4, 1918, Lieutenant Georges Painvin, a member of the Bureau, was able to identify two messages with identical strings and orderings of text, indicating plaintext with the same beginning and keys. This, and other clues led to a systematic and statistical breakdown of the German scheme during times of heavy communication. However, the Allies never did develop an universal method for decrypting ADFGVX ciphers. The largest success of Painvin was on June 3, 1918, when he intercepted and decrypted the (translated) message *"Rush munitions Stop Even by day if not seen,"* which was later confirmed, and prepared the French army for an attack on June 7 [2].

Method

The ADFGVX cipher first employs a 6×6 Polybius square to encrypt plaintext monographs into digraphs and then applies a single columnar transposition on the modified text. Simply put, the plaintext letters are substituted by digraphs and are then transposed in columns by the use of a keyword [3]. This combination of fractionation and transposition makes the cipher especially difficult for cryptanalysis [1].

Polybius Square

To accomplish the first step of encryption, plaintext characters are substituted using a 6×6 Polybius square. The Polybius square is a checkerboard scheme that uses the letters **A**, **D**, **F**, **G**, **V**, and **X**,

	A	**D**	**F**	**G**	**V**	**X**
A	O	K	Z	P	4	G
D	J	3	R	5	H	I
F	8	W	V	C	T	Y
G	1	S	D	2	E	X
V	B	M	0	A	F	L
X	7	U	N	9	Q	6

Table 1: *A sample ADFGVX Polybius Square*

in order, as column and row identifiers, which forms a grid where each element of the grid can be identified by its row and column header [3]. In this 6×6 scheme, the English alphabet and the digits 0–9 are randomly placed within the grid and each of these characters maps to an unique pair of letters. Using the sample Polybius square in Table 1, we get that the number *8* encrypts to the pair *FA*, and the pair *AD* decrypts to the letter *K*.

In the original ADFGX cipher, one uses a 5×5 Polybius square, with each letter mapping to one of the 25 different digraphs. Of course, since there are 26 letters in the English alphabet, one traditionally would merge the letters i and j into a single letter. Whereupon the correct letter would be chosen contextually when decrypting [1].

Since, in each case, a single plaintext character is mapped to more than one ciphertext character, the ADFGVX and ADFGX ciphers are examples of fractionating ciphers.

As an example, using Table 1, the plaintext message *CRYPTOGRAPHY* will result in the ciphertext:

$$FGDFF\ XAGFV\ AAAXD\ FVGAG\ DVFX. \quad (1)$$

Columnar Trasposition

To make this cipher even stronger, our already once encrypted text is encrypted again using columnar transposition, or in other words, by moving around groups of the ciphertext via some keyword.

To elaborate, suppose we have some keyword (in the language the message is written in), we eliminate all the duplicate letters of the keyword, preserving the leftmost letters. For example, if the keyword is *ZYXY* we write *ZYX*, rather than *ZXY* [1]. With the remaining ordered characters, the message is arranged so that the encoded message can be read "correctly" (that is, so that if it were decrypted it would be a readable message) from left to right, and from top to bottom. Then each character of the encrypted message is placed directly under a letter from the keyword, moving down to a new row when all letters from the keyword have been used.

As an example, using the keyword *MATHEMATICS*, which becomes *MATHEICS*, we would write out (1) as follows,

M	A	T	H	E	I	C	S
F	G	D	F	F	X	A	G
F	V	A	A	A	X	D	F
V	G	A	G	D	V	F	X

(2)

Thus the characters of the digraphs form columns underneath the letters of the keyword.

The keyword's letters are then rearranged in alphabetical order, with their corresponding columns rearranged in the same order simultaneously. In our previous example *MATHEICS* would be rearranged to read *ACEHIMST*, and (2) would become,

A	C	E	H	I	M	S	T
G	A	F	F	X	F	G	D
V	D	A	A	X	F	F	A
G	F	D	G	V	V	X	A

(3)

To complete the encryption, the left-most column is now written as a string of characters horizontally, with the topmost letters of the column becoming the leftmost in the string, the bottommost becoming the rightmost. This is then repeated for the second column, etc., until the last column [3]. This transforms (3) into the final, encrypted, string:

$$GVGAD\ FFADF\ AGXXV\ FFVGF\ XDAA. \quad (4)$$

Now the ciphertext may now be sent with some degree of confidence that it will not be decrypted by someone without a key.

Strengths

The ADFGVX cipher was designed to not only be a strong cipher, but also to be a relatively simple cipher to use. The major advantage of this cipher in the context of the First World War had to do with the substituted Polybius square. The Polybius square generally uses the digits 1–5 to act as column headers and row identifiers, which was suited for its original method of encoding (knocks or light flashes), however as mentioned above, these messages were being encoded using Morse code. Thus it is most sensible to use characters which have Morse equivalents that are all very distinct from one another, these being the letters A, D, F, G, V, and X. Examples of the International Morse symbols for the

A	· —
D	— · ·
F	· · —
G	— — ·
V	· · · —
X	— · · —

Table 2: *International Morse Symbols for ADFGVX*

letters A, D, F, G, V, and X can be seen in Table 2. By using these distinct letters, one may assume operator accuracy would increase [2].

As previously mentioned, the ADFGVX and related schemes act as a fractionating cipher since they map each character to more than one letter, as a result of the usage of the Polybius square. Additionally, the underlying concept of the Polybius square can be modified, adding additional dimensions, by the example of the ADFGVX cipher, to aid in encoding an alphabet of any length. Although, if the length of the alphabet cannot be factored into dimensions to form a Polybius square (or modified "Polybius rectangle"), then either some letters will have to map to a non-unique digraph (one letter appears in more than one position in the square), or certain digraphs will have to map back to a non-unique plaintext letter (more than one letter appears in one position of the square). Of course, this is also generalizable to greater than a digraphic scheme, in the sense that a single plaintext character could map to a string of any length by use of a higher dimensional "Polybius Hyperrectangle." However, these fractionation methods are a result of the Polybius square, and are also not traditionally used in the implementation of the ADFGVX cipher.

It is clear that the first operation of this cipher will not prevent one's ability to break the code on its own. Encryption of sample English text [4], of length 23,664 eligible characters, leads to the following sample index of coincidence,

$$IC := \sum_{i=1}^{36} \frac{n_i(n_i-1)}{N(N-1)} \simeq 0.034884 \qquad (5)$$

Which is exactly what is suspected of English text[1] (and numbers 0–9) with the approximation that the count of numbers relative to regular text is negligible. Applying the Polybius square does not change the frequency analysis performed on text if the person decrypting knows a Polybius square is being used, assuming that each character maps to only one digraph and vice-versa, since monographs are simply substituted with digraphs. For this same reason if an index of coincidence calculation was performed on the encrypted text of 36 digraphic combinations, the index of coincidence would remain the same.

It is when this method of substituting monographs with digraphs is coupled with the columnar transposition described above, that the ADFGVX cipher becomes strong [3]. By the process of columnar transposition, each of the two singular characters from every digraph is re-associated with the character above it or below it in its column, or the character at the top or bottom of the row next to it (right for top, left for bottom). In the general case, the frequency of every digraph should be changed to something that is more reflective of a polyalphabetic cipher, since two of the same digraphs may no longer correspond to the same plaintext letter. For 36 characters, the index of coincidence for a uniform distribution of text is $1/36 \simeq 0.027778$. Using Table 1 and the keyword *WARIOPNCES* to encrypt the text used to compute (5), and then computing the new index of coincidence leads to the following,

$$IC_{new} \simeq \sum_{i=1}^{36} \left(\frac{n_i}{N}\right)^2 \simeq 0.032469 \qquad (6)$$

While this change does not seem very large, we note that this is a change of about 36% towards the perfectly polyalphabetic scheme for 36 letters. A similarly sized change in the index of coincidence for the more familiar 26 letter scheme is comparable to changing the index of coincidence from 0.066700 to 0.056534. Which is the relative equivalent to changing from a monoalphabetic cipher to a polyalphabetic cipher using 2 alphabets in a 26 letter scheme [5].

In a sense, the columnar transposition allows a letter to be distributed to more than one place at a time, which hides the properties of a typical plaintext message such as letter frequency, which in turn hides the clues to how the message was transposed [2].

Weaknesses

Since the ADFGVX cipher is a private-key cipher it suffers from the weakness that if anyone has the full Polybius square and keyword, then decryption becomes a matter of writing out the message and applying the algorithm in reverse.

[1] Performing this analysis on just the text, excluding the 0–9 leads to $IC \simeq 0.067622$, which is in line with typical monoalphabetic English text as well

As mentioned prior, the application of the Polybius square alone is not very strong as it does not change any frequency analysis, since one uses strings of two characters for frequency analysis instead of singular characters. With limited observance it becomes clear that the text has been encrypted using a Polybius square. In fact, the interception of the first ADFGX messages having only five characters immediately led Painvin to suspect a checkerboard encryption scheme. This is similar for the ADFGVX cipher, except that considering the plaintext alphabet uses contains 36 characters (with the inclusion of the digits 0–9), a cryptanalyst only has to decide if the extra ten digits are to encrypt numbers or reduce frequency clues with homophones. Once one can determine that the cipher is monoalphabetic, frequency analysis can quickly solve the cipher [2]. Thus, for people without access to the keyword, breaking the ADFGVX cipher is almost entirely a matter of undoing the columnar transposition.

The addition of transposition eliminates the possibility of breaking the cipher using frequency analysis. However, once one is working under the notion of columnar transposition, breaking it can become possible given a high frequency of interceptions. Take, for instance, a case where many operators are all using the same key to encrypt their messages, and all of the messages begin with *HI THERE*. Although in the encrypted text these letters may be positioned differently, we note that when the text is arranged in columns the ciphertext that corresponds to *HI THERE* should be found in the first fourteen letters. If in the column arrangements of our messages, even one of these fourteen letters does not match the others, then there is an error in our arrangement. Thus given a large enough sample size, knowing that the messages all begin the same (or end the same), we can attempt to find the length of the keyword by identifying the matching arrangements of letters at the beginning of the message [2].

The issue with this strategy is that the messages are not required to be of a length that is divisible by the length of the keyword, and so some columns may have up to one more additional letter than another column, which can change the steps one may need to take arrangements. However, given a large enough sample size, it is likely that there will be some of a length divisible by the keyword, or only off by a small number of characters [3].

Once the length of the keyword is known, the issue becomes rearranging the columns to their original state before the transposition. Although they were a hindrance in the prior step, the varying lengths of columns becomes very useful now. Since the longer columns would have been on the left side before transposition, and the shorter ones on the right, we can concern ourselves with finding the arrangements of two subsets, which greatly decreases the number of possible arrangements. For instance, if our keyword is of length X, then there are $X!$ possible column arrangements. However, if there are Y columns of greater length, then we need only concern ourselves with $Y!(X-Y)!$ arrangements, far less than before. To test these arrangements, one carries out a frequency analysis on the adjacent pairs of letters to see if the analysis is similar to the encoded language [3]. If it is, then monoalphabetic decryption can begin to determine if the arrangement is correct. If not, then we make take further steps from the results of our analysis.

For instance, if we compare two frequency analyses performed on two different arrangements of the same message, and we find that AD is the most frequent in the first, and DA is the most frequent in the second, then it is likely that a transposition has occurred in the arrangement. To solve this issue, one need only switch around the two columns in one of the arrangements, and see if it yields a better frequency analysis on all characters [1]. By matching digraph frequency to column arrangements as such, one may slowly begin to find the correct arrangement of text.

Another technique that is effective, but only when the keyword is of even length, is to examine the frequency of a single character in the odd columns against the frequency of that same character in only the even columns. The reason for doing this is related to the way in which an even keyword arranges the digraphs into columns. Under an even keyword, the first letter of each digraph always belongs to an odd column, and the second always belongs to an even column. Using Table 1 as our example, if we examine a single character in the ADFGVX Polybius square, say A, we note that when it is the first letter in the digraph it corresponds to one of the labels: O, K, Z, P, 4, or G. Whereas if A is the second letter in the digraph, it corresponds to one of: O, J, 8, 1, B, or 7. When calculating the sum of the relative frequencies [7] of these first six letters (assuming the occurrence of digits is approximately zero), we obtain a value of 0.12297, while the sum of the relative frequencies of the other six letters is 0.09152, two very different frequencies. This difference should occur in general for any character we choose, since it is likely that six random characters of the Polybius square will have very different relative frequency from the

other six (with only one letter in common). Thus, when we test even and odd columns in this manner, we expect the frequency analysis to be quite different between even and odd columns. If not, then we may suspect that an even column is arranged as an odd column (and vice-versa) [6].

Using all of these techniques in a large sample size should exhaust many of the potential possibilities, eventually leading to the key of the cipher, or a key of the same length and ordering. In essence, the solution to an ADFGVX cipher is a combination of brute-force and statistical analysis, as statistical analysis is used to: guess the length of the keyword, find the original arrangement of columns, and then decrypt the monoalphabetic fractionated text.

A Sample Passage

Using the Polybius Square from Table 1 and the keyword *WARRIOR PRINCESS*, which becomes *WARIOPNCES*, we translate the plaintext [8]:

```
HERE THE RED QUEEN BEGAN AGAIN CAN
YOU ANSWER USEFUL QUESTIONS SHE
SAID HOW IS BREAD MADE I KNOW THAT
ALICE CRIED EAGERLY YOU TAKE SOME
FLOUR WHERE DO YOU PICK THE FLOWER
THE WHITE QUEEN ASKED IN A GARDEN
OR IN THE HEDGES WELL IT ISNT
PICKED AT ALL ALICE EXPLAINED ITS
GROUND HOW MANY ACRES OF GROUND
SAID THE WHITE QUEEN YOU MUSTNT
LEAVE OUT SO MANY THINGS FAN HER
HEAD THE RED QUEEN ANXIOUSLY
INTERRUPTED SHELL BE FEVERISH AFTER
SO MUCH THINKING SO THEY SET TO
WORK AND FANNED HER WITH BUNCHES OF
LEAVES TILL SHE HAD TO BEG THEM TO
LEAVE OFF IT BLEW HER HAIR ABOUT SO
```

Which becomes the, doubly-long, ciphertext:

```
VVVVV GFXFF VDXDG AAFVA GXFVF ADDAV
FDDVV VVVDX GFFVV XXGVG XVGFX FDXVX
FXVVV ADXVD FXVVG VVFXX VDFXV FGFDV
XAXVF FFVVF DVDXV AVVGV AVGAD GDXVX
DXVGG XFGGD GVGXD VFGGF VAVFG AFGFD
DXVGA GDGAG FFAFV GVGGX AXDVX VDDXF
XFFFV DGDGD GGGGX DGGGG DGVFA GGFAX
XDDXG VGVGF AVGAF FDVFD GXXVV VXFGV
GXGGD GVAFV GGAFF GVDDF DDAFD XXGVG
DDGFD XFVVF AXFAD VFAAG FDXXV XVXVF
AXDFG GXXDD FDVFG FADXA DGFVV GFVDV
FDDGG GVGDG DDXVV FVXXG ADDDV AVXDF
DXDVG XGXDV VDVAG VXFDV VDFFA VVDXD
XFXXX XXFFD GDFFF DVVDV VAAXF GFFFD
GAXVG DXVXV DGXFV DADVV FVGAG VDGAV
AFVXV XAAFV VFGGG DDDXD FVFXV GDVGV
VXXVD VFFXX VAVVV FVGFF VFDXF GGXGG
FVAVG GAADV VVFDF GDDVX FVVVV FDXFV
VAFDV AVFXV VDGGF DAVVF XVFVV XVXVF
FDDFG GGVXF XDVXD GVAVG GAVDD GDAAV
ADGXA VGGFG GDVXX DGVDA GVVGV GAXFF
GAXDG FGAGV GGXDV GVGXX DVFXG VVDVD
XDDAF FDGXG FXGGG VDAFF VVVDV VGVFD
AFXFG VVVVD DVDGF FVXGF VXGAX DVAGV
DVXDG FXVXV XVVVD VGVXV DDAFF VDGVX
DXDVF VGVDV VVFVX DFFVV VVVVD VDDVV
AFFVX DVXVX VVXDV AVDVG VGDGG AAFAG
XXGAD DFDVD FFFDV VXGGX GAAFV DFGGA
XDAFG GVGDG VDGDD DGVVG DGGFG GXFGA
AFXVD GDXVA FGAGV GDVGF DXFGA AVGDD
FAVFG VVDAF VVGDA A
```

References

[1] Greg Goebel, *Codes & Codebreakers In World War I* http://www.vectorsite.net/ttcode_04.html#m3 (2014).

[2] David Khan, *The Codebreakers – The Story of Secret Writing*, Macmillan Publishing Company, New York, 1st edition, (1967).

[3] Cecily Morrison, and Ben Roberts, *Codes and Ciphers - ADFGVX Cipher*, http://www.srcf.ucam.org/~bgr25/cipher/adfgvx.php (2008).

[4] James de Mille, *A Strange Manuscript Found in a Copper Cylinder*, Harper & Brothers, New York, (1888).

[5] *dCode*, http://www.dcode.fr/ (2016).

[6] James Lyons, *Practical Cryptography*, http://practicalcryptography.com/ciphers/adfgvx-cipher/ (2012).

[7] Robert Edward Lewand, *Cryptological Mathematics*, The Mathematical Association of America, (2000).

[8] Lewis Carroll, *Through the Looking-Glass*, Macmillan Publishers, (1871).