Cryptodiagnosis of "Kryptos K4"

Richard Bean
School of Information Technology and Electrical Engineering
University of Queensland, Australia 4072
r.bean1@uq.edu.au

Abstract

The sculpture "Kryptos" at the Central Intelligence Agency in Langley, Virginia, contains four encrypted passages. The last, known as "K4" and consisting of 97 letters, remains unsolved.

In this work, we look at unusual statistical properties of the K4 ciphertext, together with the known plaintext, using Monte Carlo sampling to perform permutation testing. This provides evidence strongly indicating a definite "one-to-one" relationship between corresponding plaintext and ciphertext letters. It also points toward a possible encryption method which could account for most or all of the observed properties. This is the "Gromark" cipher invented by Hall (1969, 1975) and analyzed by Blackman (1989).

1 Introduction

The "Kryptos" sculpture was installed at the Central Intelligence Agency (CIA) in Langley, Virginia in November 1990. The sculptor was Jim Sanborn and the cryptographic consultant was Ed Scheidt, who retired from the CIA in December 1989. The sculpture contains four encrypted messages totalling 865 letters plus 4 question marks.

Scheidt has indicated that the codes were designed to be solved in five, seven or ten years.

The first three sections were solved independently by three different teams or individuals: an NSA team in 1992, David Stein from the CIA in 1998, and Jim Gillogly in 1999. The fourth section, "K4", consisting of 97 letters remains unsolved and its encryption method remains publicly unknown. During the period 2010 to 2020, four parts of the K4 plaintext with locations were released by the sculptor, totalling 24 letters. Further details may be found in Dunin and Schmeh (2020).

Callimahos (1977) and Lewis (1992) describe the process of diagnosis of an unknown cipher type. Callimahos, in a chapter entitled "Principles of Cryptodiagnosis", sets out a process of hypothesis formulation and hypothesis testing. This involves arrangement and rearrangement of the data to disclose nonrandom characteristics, followed by recognition and explanation of these characteristics. The chapter headers are: manipulating the data, recognizing the phenomena, and interpreting the phenomena.

Lewis states that the task of an analyst is finding, measuring, explaining, and exploiting a phenomenon (or phenomena). Writing about cipher type diagnosis, he describes the search for "something funny" or "finding the phenomena".

Since these publications, Mason (2012) prepared a table of cipher statistics for many American Cryptogram Association (ACA) types, with associated random forest (2013) and neural net (2016) classifiers. Nuhn and Knight (2014) also developed a classifier for ACA cipher types using a support vector machine approach.

In this paper we attempt to measure some of the interesting phenomena seen in K4 and provide possible explanations. We perform statistical testing using Monte Carlo sampling and describe one possible encryption method, the Gromark cipher of the ACA, and its variants. Finally we conduct an extensive search of the Gromark key space for various bases and key primer lengths before discussing our conclusions.

2 Analysis

Good (1983) commented on the practice of looking at a sample ciphertext and deciding on a test of significance based on the observed data, instead of running a standard series of tests. The passage is worth quoting in full to describe the risks and rewards of such an approach.

... it is sometimes sensible to decide on a significance test after looking at a sample. As I've said elsewhere this practice is dangerous, useful, and often done. It is especially useful in cryptanalysis, but one needs good detached judgment to estimate the initial probability of a hypothesis that is suggested by the data. Cryptanalysts even invented a special name for a very far-fetched hypothesis formulated after looking at the data, namely a "kinkus" (plural: "kinkera"). It is not easy to judge the prior probability of a kinkus after it has been observed.

2.1 Ciphertext analysis

One of K4's most prominent unusual features is the number of repeated bigrams when the ciphertext is written at width 21 (Hannon, 2010; LaTurner, 2016; Kirchner, 2003); see Table 1.

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

Table 1: K4 ciphertext written at width 21

If we consider the 76 bigrams formed vertically (starting with OF and finishing with GR), there are 11 repeated bigrams (AZ BS IT LS LW PK QZ SN WA ZT KK). This value is in line with the expected number of repeated bigrams if a typical English plaintext was written out at width 21; for example, testing all 97-letter contiguous subsets of the King James Bible gives an average value of 9.7 repeated bigrams at width 21.

If we perform Monte Carlo sampling and take a large number of permutations of the ciphertext (Good, 2013), we can estimate the proportion of permutations of the ciphertext which would have at least this number of repeated bigrams. In this case, this proportion is approximately one in 6,750. Programs written in C to calculate values in this paper are provided via Github.[1]

The recently solved (Oranchak et al., 2020) "Zodiac 340" cipher also had a similar property (Daikon, 2015; Van Eycke, 2015). A relatively high number of repeated bigrams was seen at width 19 in the ciphertext. The cipher was ultimately found to be a combination of transposition and homophonic substitution. The width 19 property can thus, after the fact, be deemed "causal" as the enciphering process caused this property to appear.

2.2 Seriated ciphers

The "seriated Playfair" cipher of the ACA might provide a partial aesthetic explanation for the width 21 patterns. This cipher is digraphic and works by performing Playfair encryption on vertical pairs of letters. That is, any given pair of letters in plaintext (p_1, p_2) maps to another pair of letters (c_1, c_2) in a one-to-one fashion. Thus, numbering the positions 0 to 96, the repeated "BS" bigrams at positions 12/33 and 18/39 would reflect the same underlying plaintext, or "AZ" at positions 49/70 and 57/78. Similarly, the "double box cipher" or Doppelkastenschlüssel, sometimes referred to as "double Playfair", described by David (1996) is a digraphic cipher which required seriation at width 21.

There are also several arguments against this as a method:

- according to the "ACA Cipher Statistics" webpage of Mason (2012), the index of coincidence (IC) of a typical "seriated Playfair" ciphertext is 0.048 with standard deviation 0.003 versus K4's IC of 0.036
- 26 letters occur in the ciphertext, while the most common Playfair variant uses only 25 from a 5x5 square
- the doublet "KK" is present, which cannot occur in standard Playfair
- a plain interpretation is that there are 97 ciphertext letters, an odd number, while Playfair works on pairs of letters. As 97 is also prime, this is also an argument against the Hill cipher suggestion of Bauer et al (2016).

The original description of the Playfair cipher by Wheatstone entitled "Specimen of a Rectangular Cipher", seen in Kahn (1996, p. 199) uses

[1] https://github.com/RichardBean/k4testing

a 9x3 rectangle, and enciphers doubled letters, which would account for the last three observations. We could take the question mark before "OBKR" on the sculpture as the initial character, with 27 different characters and 98 ciphertext characters. The low IC could then be accounted for by careful selection of the key.

However, these theories all became moot after the "BERLIN" plaintext clue was released in November 2010. This corresponds to the ciphertext "NYPVTT". Thus, if a seriated digraphic cipher at width 21 had been used to encipher the plaintext, we would have two different plaintext bigrams ending in "I" and "N" both mapping to "ZT", which is impossible. As the letter "K" in the 2014 plaintext clue of "CLOCK" also enciphered to "K" this ruled out the use of standard Playfair for the vertical bigrams.

We might also wish to check a width of seven, based on a purely aesthetic argument, since 98 characters is seven pairs of rows with seven characters per row. Similarly, the "NORTHEAST" plaintext clue was released in January 2020, which corresponded to letters 26-34 in the ciphertext, "QQPRNGKSS". If a seriated digraphic cipher had been used to encipher the plaintext at a width of seven, we would have two different plaintext bigrams ending in "N" and "O" both mapping to "BQ", again impossible.

2.3 Other observations

Many other statistical anomalies have been noted by others. Previous solvers of Kryptos have noted the repeated doublets (BB, QQ, SS, ZZ and TT) in the same columns when the ciphertext is written at width seven. These letters are shown in bold in Table 1. An NSA analyst (Redacted, 2007) and Gillogly (Gillogly, 1999a) suggested this property could be due to combined polyalphabetic substitution and transposition. The width 21 property could also be used to argue for combined transposition and substitution, as with the Zodiac 340 cipher; however, this paper argues against a transposition step.

Stehle (2000) noted that the ciphertext segment "DIAWINFBN" has the property that when converted to numbers (from the standard alphabet), 0 to 25, the difference between the first five letters and the corresponding letters four positions right is 5 (modulo 26). Thus *I* minus *D* corresponds to 8 minus 3, *N* minus *I* to 13 minus 8, and so on.

These observations are unusual, and may well be causal, but were ultimately considered harder to measure, explain or exploit than the observations discussed here.

2.4 Known plaintext analysis

The 24 known plaintext letters are as follows: "FLRVQQPRNGKSS" in cipher corresponds to "EASTNORTHEAST" in plain and "NYPVTTMZFPK" in cipher corresponds to "BERLINCLOCK" in plain.

Materna (2020) noted that for the known K4 plaintext, where the plaintext letters are in the set $\{K,R,Y,P,T,O,S\}$ the corresponding ciphertext letters are very close in the standard alphabet to the plaintext letters. Thus, the 10 shortest distances (the so-called "minor differences" (Friedman, 1954)) sum to 21, as shown in Table 2, for a mean of 2.1.

Plaintext letter	S	T	O	R	T	S	T	R	O	K
Ciphertext letter	R	V	Q	P	R	S	S	P	F	K
Distance	1	2	2	2	2	0	1	2	9	0

Table 2: Minor differences between plain and ciphertext letters

Monte Carlo sampling by permuting the ciphertext letters of K4 demonstrates this occurs only in about one in 5,520 permutations of K4 ciphertext letters.

With the release of 24 known plaintext characters, we can create a table showing, for each repeated plaintext letter, what the corresponding ciphertext letter set is, and then measure the shortest distance between each of the ciphertext letters. The repeated known plaintext letters are A, C, E, L, N, O, R, S and T. Table 3 measures the minor differences between the ciphertext letters corresponding to each, producing 13 values.

We note that the mean is 3.6 and 10 of 13 values are less than five. Performing Monte Carlo sampling and permuting the ciphertext randomly, we found that about one in 240 permutations have a mean less than or equal to 3.6, while about one in 310 permutations have at least 10 values less than five.[2]

[2]Looking at the distances in the Kryptos alphabet, 7 of 13 values are multiples of 5. If the plaintext letters are numbered 0-25 from the standard alphabet and the ciphertext letters are numbered 0-25 from the Kryptos alphabet reversed, then the sequence plain minus cipher modulo 26 is calculated, 13 of 24 values are multiples of 5; randomly permuting the cipher-

Plain	Cipher	Distances in standard alphabet	Distances in "KRYPTOS" alphabet
A	KL	1	9
C	MP	3	11
E	FGY	1,7,8	1,10,11
L	VZ	4	3
N	QT	3	10
O	QF	11	8
R	PP	0	0
S	RS	1	5
T	RSV	1,3,4	5,5,10

Table 3: K4 distances between cipher letters corresponding to repeated plaintext letters

These observations are unusual and strongly suggest that "one-to-one" encryption of single letters to single letters is occurring; that is, there is no transposition involved in the encryption process. It is of course possible that a new encryption algorithm never before seen is in use, but this would render solution very unlikely.

2.5 The Gromark Cipher

Instead, we suggest that these observations are most compatible with the "Gromark" cipher (Hall, 1969; Rogot, 1975a; American Cryptogram Association, 2016). This was also a suggestion of Gillogly (1999a; 1999b; 2004b; 2004a) before known plaintext was made available in 2010.

Gromark as described by Hall (1969) operates by using a "primer" of five digits, which is expanded to form a key of the length of the plaintext, using a "lagged Fibonacci generator" by continually adding the first two available digits, starting with and including the primer, to get the next key digit (modulo 10).

A plain and cipher alphabet are used; in ACA puzzles, the standard alphabet is used for the plain (A to Z) and the primer is given. The plain and cipher alphabets are written in rows with the plain on the top. The key digit corresponding to a particular plaintext letter is then used to count that many steps right from the corresponding letter in the cipher alphabet to produce each ciphertext letter.

Rogot (1975b) pointed out that, analogously to the "Quagmire" cipher types, various kinds of plain and cipher alphabets can be used. They can be standard or keyword-based. Lewis (1992, p. 116) wrote about using the same alphabet for plain and cipher, or using a superadditive numeric key.

Thus, one explanation for the "minor differences" observations in the "Analysis" section above could be that the cipher alphabet is "near" the standard A-Z alphabet, perhaps based on a keyword, and then the minor differences between ciphertext letters corresponding to repeated plaintext letters are small numbers.

Blackman (1989) considered further variations, such as using a non-decimal base, varying the length of the primer, or as in Barker (1984), using a different rule for building up the key.

Holden (2018) used Gromark as an illustration of the concept of the "linear feedback shift register" (LFSR) which is more fully described in Barker (1984).

Rogot (1975a), Deavours (1977a) and Blackman (1989) all noted that with an even base and a five digit primer, there is an underlying structure of length 21 in the key and ciphertext. For example, Deavours remarked that writing such Gromark ciphertext out at width 21, each column is encrypted by either all even or all odd key digits, and with enough ciphertext, the underlying structure of the primer is revealed. Blackman extended this approach to recovery of the base and length of the primer by examination of the index of coincidence, although typically a ciphertext of length much greater than 100 letters is required.

The more general concept of inferring a sequence generated by a pseudo-random number generator (PRNG)[3] from known terms is dealt with in more detail in Reeds (1977), Plumstead (1982), Knuth (1985), and Boyar (1989). For instance, Boyar wrote about a linear congruential recurrence with n terms:

$X_i = a_1X_{i-1} + a_2X_{i-2} + \ldots + a_nX_{i-n} + a_{n+1} \pmod{m}$

In the case of the standard ACA Gromark cipher, we have $m = 10$, $n = 5$, $a_4 = a_5 = 1$, and $a_1 = a_2 = a_3 = a_6 = 0$.

If a base two, five digit Gromark cipher is used, with standard English plaintext taken from the King James Bible, simulations indicate that for any given key, about one in 10 ciphertexts will have the property of at least 11 repeated vertical

text, this is a 1 in 1,470 result. This might imply a method involving 5x5 Polybius squares, such as a conjugated matrix bifid; but nothing was found.

[3] It is curious, but probably not relevant, that this is the only 4-letter sequence occurring twice in the ciphertext part of the sculpture.

bigrams at width 21. By way of explanation, Table 4 shows the key expansion beginning with the primer 00001. Ten of the values in each complete row are the digit 1, and the vertical bigrams are enciphered with either 00 or 11; thus, enciphering will tend to preserve existing patterns of vertical bigrams present in the plaintext.

0 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 1 1 1 1 1
0 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 1 1 1 1 1
0 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 1 1 1 1 1
0 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 1 1 1 1 1
0 0 0 0 1 0 0 0 1 1 0 0 1

Table 4: Gromark binary key

This "one in 10" proportion is very different from the "one in 6,750" result obtained from the Monte Carlo sampling above. Similarly, for a particular plaintext and sufficiently large base, it is generally simple to find a five digit primer which results in the ciphertext having the property of a large number of repeated vertical bigrams.

The known K4 plaintext now indicates the base must be at least three, because some plaintext letters encipher to at least three different ciphertext letters.

Additional arguments for the use of the Gromark cipher include:

- Gromark was described by Blackman (1989) as a "pencil-and-paper field cipher". Similarly, Scheidt has been quoted as stating: "K4 cryptography is similar to what would be provided agents or pilots in case of capture" (Hannon, 2011);
- Gromark is definitely "more than one stage" as the primer must be expanded to the complete key. Scheidt stated in 2015 that "[he] would consider [K4 encryption] [to be] more than one stage". (Schmeh, 2015);
- Gromark does not involve transposition and enciphers letters to letters. Sanborn has been quoted as stating: "BERLINCLOCK in plain matches directly with NYPVTTMZFPK. It is a one-to-one match with plain B taken, has the encipherment done to it, and out pops a cipher N, plain E is then enciphered to a cipher Y" (Bogart, 2019);
- Gromark is one of the few ACA ciphers in Mason's table to result in a "flat" index of coincidence, that is, one close to the value 1/26 = 0.03846. The IC value of K4 is 336/97/96 = 0.03608;
- The unique feature of the Berlin Clock is that it uses base 5 or 12 arithmetic (Schridde, 2020) and Sanborn has stated "you'd better delve into that particular clock" (Schwartz, 2014);
- Scheidt hints about base arithmetic in the 2015 interview above and also in 2020: "if you can change the language base then it becomes in my favor and not your favor of trying to break it. It becomes more of a challenge now, when it was used as the mask it was current, 2020 secret." (Jacobs, 2020). This may refer to Blackman (1989);
- The raised letters on the sculpture stylized as "$D^{YA}H^{R}$" may refer to a Gromark five-digit primer and are reminiscent of binary.[4] Indeed, the "Vimark" cipher (Dickerhoof, 1971) is just Gromark at base 26 using numeric values of letters.

Arguments against use of the Gromark cipher are:

- The initial ACA experience showed that Gromark encryption is error-prone and all ACA challenges are now provided with a check digit. However, the most error-prone aspect is the key expansion stage; this could be checked by a third party without revealing the plaintext.
- Sanborn has stated that he is an "anathemath" on several occasions (Allsop, 2010);
- At the 2011 "Kryptos Dinner" at the Zola Restaurant in Washington DC, Scheidt stated "[K4 cryptography] is not mathematical (although this does not preclude it being modeled mathematically), it is simple, can be remembered, and executed years later when used with the correct key word/s." (Hannon, 2011)

[4] Alternatively, this may be a reference to historical codes. Telegraph and telex messages were charged per word sent. To reduce costs, large international companies (mostly banks) developed and used five letter codes. Codes such as Acme had error correction features which in time were replaced by binary error correction systems.

- Typically, the ACA version of the cipher uses a "standard" plain alphabet, that is, A-Z in order. With the release of the "CLOCK" crib, C and R in the plaintext alphabet would both need to map to P in the ciphertext alphabet but are more than 10 places apart in the standard alphabet.

- Assuming Gromark is in use, there is a tension between the observations concerning the IC, the ciphertext bigrams at width 21, and the base chosen. A low base means that the number of different vertical bigrams in the key may be low; but on the other hand, very few ciphertext outputs will have an IC as low as 0.036. If, however, the encipherer wants to deliberately insert the width 21 property, with a higher base, they have many primers to choose from to achieve that property.

The given plaintext maps plain T to cipher V at position 24, and L to V at position 66 (numbering the positions 0 to 96). If Gromark was used as the cipher, this implies the key is not repeating at period 21 or 42. Perhaps the period is 63 or 84. A period of 63 is reminiscent of an *m*-sequence or "maximal length sequence" as seen in a particular example of Golomb and Gong (2005). In this example, Golomb and Gong produced an *m*-sequence over $\mathbb{F}_{2^2}$ of degree 3 with period 63 using the irreducible polynomial x^2+x+1.

This sequence (extended to 84 entries) is shown in Table 5, with the field elements in $\mathbb{F}_{2^2}$ $\{0,1,\beta,\beta^2\}$ replaced by $\{0,1,2,3\}$. As with the binary Gromark key, the number of distinct vertical bigrams is quite low; only four: $00, 12, 23$ and 31.

1	1	1	0	1	3	0	0	1	1	2	3	0	1	0	3	1	3	1	1	3
2	2	2	0	2	1	0	0	2	2	3	1	0	2	0	1	2	1	2	2	1
3	3	3	0	3	2	0	0	3	3	1	2	0	3	0	2	3	2	3	3	2
1	1	1	0	1	3	0	0	1	1	2	3	0	1	0	3	1	3	1	1	3

Table 5: Golomb and Gong *m*-sequence of period 63

Meanwhile, a period of 84 is often seen with base eight and primer length five, further explained below.

3 Search

Given the 24 known plaintext letters, we discovered that a simulated annealing search (see for instance Lasry (2018) using hexagram statistics for scoring as in Bean (2020)) for the plain and cipher alphabets would eventually converge for a given key, when the alphabets were allowed to vary.

A set of inequalities and equalities was developed to narrow down the possible primers for base and primer length possibilities. By means of this reduction, the entire search space for base 10, length 5 was examined thoroughly.

If we number the numeric key from 0 to 96, so that each key digit is denoted by $k_0,\ldots,k_{96}$, we can write out the relationships between the 24 known plaintext and ciphertext letters. The "p" and "c" functions here calculate the numerical equivalent of a given letter in the plaintext and ciphertext alphabets (0 to 25). Then, pairs of these relationships imply relationships between digits of the key.

- $p(T)+k_{24}=c(V), p(T)+k_{28}=c(R), p(T)+k_{33}=c(S) \implies k_{24}\neq k_{28}, k_{28}\neq k_{33}, k_{24}\neq k_{33}$

- $p(E)+k_{21}=c(F), p(E)+k_{30}=c(G), p(E)+k_{64}=c(Y) \implies k_{21}\neq k_{30}, k_{21}\neq k_{64}, k_{30}\neq k_{64}$

- $p(R)+k_{27}=c(P), p(R)+k_{65}=c(P) \implies k_{27}=k_{65}$

- $p(N)+k_{68}=c(T), p(N)+k_{25}=c(Q) \implies k_{68}\neq k_{25}$

- $p(A)+k_{22}=c(L), p(A)+k_{31}=c(K) \implies k_{22}\neq k_{31}$

- $p(L)+k_{66}=c(Q), p(L)+k_{70}=c(Z) \implies k_{66}\neq k_{70}$

- $p(O)+k_{26}=c(Q), p(O)+k_{71}=c(F) \implies k_{26}\neq k_{71}$

- $p(C)+k_{69}=c(M), p(C)+k_{72}=c(P) \implies k_{69}\neq k_{72}$

- $p(S)+k_{23}=c(R), p(S)+k_{32}=c(S) \implies k_{23}\neq k_{32}$

- $p(O)+k_{71}=p(E)+k_{21}=c(F) \implies k_{71}\neq k_{21}$

- $p(N)+k_{25}=p(O)+k_{26}=c(Q) \implies k_{25}\neq k_{26}$

- $p(T)+k_{24}=p(L)+k_{66}=c(V) \implies k_{24}\neq k_{66}$

- $p(A)+k_{31} = p(K)+k_{73} = c(K) \implies k_{31} \neq k_{73}$
- $p(H)+k_{29} = p(B)+k_{63} = c(N) \implies k_{29} \neq k_{63}$
- $p(S)+k_{32} = p(T)+k_{33} = c(S) \implies k_{32} \neq k_{33}$
- $p(I)+k_{67} = p(N)+k_{68} = c(T) \implies k_{67} \neq k_{68}$
- $p(R)+k_{27} = p(C)+k_{72} = c(P) \implies k_{27} \neq k_{72}$
- $p(S)+k_{23} = p(T)+k_{28} = c(R) \implies k_{23} \neq k_{28}$

For base 10, primer length 5, out of the initial 99,999 possible non-zero keys, this left 1,040 remaining. If the digits in each key were randomly chosen and uncorrelated within each key, we have 21 inequalities and one equality at base 10; the proportion of keys satisfying all these would be $(\frac{9}{10})^{21}(\frac{1}{10})$ = 0.01094, which implies in some sense that the Gromark key digits are approximately "random".

After this, we can apply further restrictions. The SageMath software (Stein, 2007) allows us to compute the Gröbner basis for the set of equations showing the relationship between the 24 plaintext and ciphertext letters. This leads to another set of 14 inequalities and one equality which each have either four or six terms. The full set may be found in the Github source.

This process ultimately showed that only 39 different primers (for the base 10 five digit case) could lead to the 24 letters of known plaintext in the correct positions.

Two of these primers, 26717 and 84393, are equivalent, up to length 97, using a variation of an observation of Blackman (1989): the keys are inverses of each other (modulo 10). So, for any given plain and cipher alphabet P and C, the result of encrypting with 26717 is equal to the result after encrypting with 84393, with the original alphabets P and C reversed. See Table 6.

These are the only two of the 39 keys which use only nine different digits. Of the 99,999 keys of length 97, only 88 keys do not contain the zero digit anywhere.

One of Blackman's ideas is that, for a given numeric key, the index of coincidence can be calculated for the ciphertext letters corresponding to each digit, and the average taken. This is the method used to determine the most likely key primers. In this case, starting with the key 98800 gives an index of coincidence of 0.0625, which is the highest of any of the keys and closest to the index of coincidence of typical English plaintext.

2 6 7 1 7 8 3 8 8 5 1 1 6 3 6 2 7 9 9 8 9
6 8 7 7 5 4 5 4 2 9 9 9 6 1 8 8 5 7 9 6 3
2 6 5 9 5 8 1 4 4 3 9 5 8 7 2 4 3 5 9 6 7
8 4 5 3 5 2 9 8 8 7 1 7 6 5 8 8 3 1 3 6 1
4 4 9 7 5 8 3 6 2 3 1 9 8 5

Table 6: Expansion of 26717 primer at base 10

The restrictions above can be applied to primers of other bases and key lengths: for instance, the only possible base 10, four digit primers are 3301[5], 6740, and 9903, and the four possible base eight, five digit primers include 00351 and 00537. As seen in Table 7 the expansions of these base eight keys have period 84 and the extra property that all columns at width seven, as well as at width 21, have either all odd or all even key digits.

0 0 3 5 1 0 3	0 6 1 3 3 6 7	4 6 1 5 3 2 7
6 0 5 1 5 6 5	6 6 3 3 3 4 1	6 6 7 5 7 4 5
4 4 3 1 1 0 7	4 2 1 7 3 6 3	0 2 1 1 3 2 3
2 4 5 5 5 6 1	2 2 3 7 3 4 5	2 2 7 1 7 4 1
0 0 3 5 1 0 3	0 6 1 3 3 6 7	

Table 7: Expansion of 00351 primer at base eight

After this, different key expansion rules can be tried, perhaps inspired by the raised letters on the sculpture. We restricted ourselves to rules where the first digit in the primer (shift register) is used in the generation function, as explained in Beker and Piper (1982, p. 183).

Although many plaintexts close to ordinary English were discovered, none were entirely convincing. If a Gromark variant was indeed used in the K4 encryption process, with a more general key expansion rule, it becomes difficult to test all the possibilities. Instead, it may be worth considering implications of the other observations in this paper.

4 Conclusion

With the use of Monte Carlo sampling analysis, the known plaintext released by Sanborn pro-

[5]Which reminds one of the Internet mystery "Cicada 3301"

vides strong indications that transposition is not involved in the K4 encryption process.

If the "Gromark" cipher of the ACA was used as the encryption method, this would explain many of the observed properties of the ciphertext and known plaintext. The "unicity distance" (Deavours, 1977b) of the Gromark cipher is approximately 48 letters, not accounting for the numeric primer, which means the solution would be unique at a ciphertext length of 97 letters.

As the Gromark cipher is the inspiration for another high-security cipher of Rubin (1996) such a cipher may be quite difficult to solve, fulfilling Sanborn's stated intention of it "going on for a century, hopefully long after my death." (Sanborn, 2009)

Acknowledgments

The author wishes to thank Ed Hannon for his extensive correspondence on this subject, and also Jim Gillogly and Eleanor Joyner (SCRYER and HONEYBEE of the ACA).

References

Laura Allsop. 2010. Kryptos sculpture inspires hope in weary code-breakers. http://edition.cnn.com/2010/SHOWBIZ/11/26/CIA.sculpture.clue/index.html.

American Cryptogram Association. 2016. Gromark: Gronsfeld with mixed alphabet and running key. https://www.cryptogram.org/downloads/aca.info/ciphers/Gromark.pdf.

Wayne G Barker. 1984. *Cryptanalysis of Shift-Register Generated Stream Cipher Systems*, volume 39. Aegean Park Press.

Craig Bauer, Gregory Link, and Dante Molle. 2016. James Sanborn's Kryptos and the matrix encryption conjecture. *Cryptologia*, 40(6):541–552.

Richard Bean. 2020. The use of Project Gutenberg and hexagram statistics to help solve famous unsolved ciphers. In *Proceedings of the 3rd International Conference on Historical Cryptology HistoCrypt 2020*, number 171, pages 31–35. Linköping University Electronic Press.

Henry Beker and Fred Piper. 1982. *Cipher systems: the protection of communications*. Northwood Books, London.

Deane R Blackman. 1989. The Gromark Cipher, and Some Relatives. *Cryptologia*, 13(3):273–282.

Bob Bogart. 2019. CNN Documentary about Kryptos a making of report with many photographs. https://scienceblogs.de/klausis-krypto-kolumne/2019/03/15/cnn-documentary-about-kryptos-a-making-of-report-with-many-photographs/.

Joan Boyar. 1989. Inferring sequences produced by pseudo-random number generators. *Journal of the ACM (JACM)*, 36(1):129–141.

Lambros D. Callimahos. 1977. *Military Cryptanalytics Part III*. National Security Agency. https://www.governmentattic.org/39docs/NSAmilitaryCryptalyticsPt3_1977.pdf.

Daikon. 2015. Things I noticed about Z340. http://www.zodiackillersite.com/viewtopic.php?f=81&t=2625.

Charles David. 1996. A World War II German army field cipher and how we broke it. *Cryptologia*, 20(1):55–76.

Cipher A Deavours. 1977a. The kappa test. *Cryptologia*, 1(3):223–231.

Cipher A Deavours. 1977b. Unicity points in cryptanalysis. *Cryptologia*, 1(1):46–68.

Dean W. Dickerhoof. 1971. The Vimark Cipher. *The Cryptogram*, 37(4).

Elonka Dunin and Klaus Schmeh. 2020. *Codebreaking: A Practical Guide*. Hachette UK.

William F. Friedman. 1954. Basic cryptologic glossary. https://www.nsa.gov/Portals/70/documents/news-features/declassified-documents/friedman-documents/publications/FOLDER_234/41761109080025.pdf.

Jim Gillogly. 1999a. another news article on Kryptos. https://groups.google.com/g/sci.crypt/c/HTQqcW9XDAI/m/yC_JQZYxBPUJ.

Jim Gillogly. 1999b. Kryptos Morse code. https://groups.google.com/g/sci.crypt/c/d3SNKxTYsBA/m/BoQwYMWdWQIJ.

Jim Gillogly. 2004a. Non-periodic polyalphabetic substitutions. http://kryptos.yak.net/63.

Jim Gillogly. 2004b. re: new member. https://kryptos.groups.io/g/main/message/1236.

Solomon W Golomb and Guang Gong, 2005. *Signal design for good correlation: for wireless communication, cryptography, and radar*, pages 134–135. Cambridge University Press.

Irving John Good. 1983. *Good thinking: The foundations of probability and its applications*. U of Minnesota Press.

Phillip Good. 2013. *Permutation tests: a practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

WJ Hall. 1969. The Gromark cipher (Part 1). *The Cryptogram*, 35(2):25.

Edward Hannon. 2010. Novel K4 Results. https://kryptos.groups.io/g/main/message/10213.

Edward Hannon. 2011. Oct 8th DC Meet with Sanborn and Scheidt. https://kryptos.groups.io/g/main/message/12611.

Joshua Holden. 2018. *The mathematics of secrets: cryptography from Caesar ciphers to digital encryption*. Princeton University Press.

AJ Jacobs. 2020. Ed Scheidt Kryptos Transcript. https://kryptos.groups.io/g/main/files/PersonalFolders/AJJacobs/EdScheidtKryptosTranscript.pdf.

David Kahn. 1996. *The Codebreakers: The comprehensive history of secret communication from ancient times to the internet*. Simon and Schuster.

Tim Kirchner. 2003. Don't know what this means, but... . https://kryptos.groups.io/g/main/message/232.

Donald Knuth. 1985. Deciphering a linear congruential encryption. *IEEE Transactions on Information Theory*, 31(1):49–52.

George Lasry. 2018. *A methodology for the cryptanalysis of classical ciphers with search metaheuristics*. kassel university press GmbH.

Geoffrey LaTurner. 2016. New member introductory message. https://kryptos.groups.io/g/main/message/18246.

Frank W. Lewis. 1992. *Solving Cipher Problems: Cryptanalysis, Probabilities and Diagnostics*. Aegean Park Press, Laguna Hills, CA.

William Mason. 2012. ACA Reference Statistics. https://bionsgadgets.appspot.com/gadget_forms/acarefstats.html.

William Mason. 2013. Compare unknown cipher against ACA cipher types (extended). http://bionsgadgets.appspot.com/gadget_forms/refscore_extended.html.

William Mason. 2016. Neural net ID test collection. http://bionsgadgets.appspot.com/gadget_forms/nnet_id_test_collection.html.

Greg Materna. 2020. Keyword for K4. https://aivirai.com/2020/08/24/4-3-kryptos-aivirai-muko-series-and-the-keyword-for-k4/.

Malte Nuhn and Kevin Knight. 2014. Cipher type detection. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1769–1773.

David Oranchak, Sam Blake, and Jarl van Eycke. 2020. Z340 has been solved! http://www.zodiackillersite.com/viewtopic.php?f=23&t=5079.

Joan B Plumstead. 1982. Inferring a sequence generated by a linear congruence. In *23rd Annual Symposium on Foundations of Computer Science (sfcs 1982)*, pages 153–159. IEEE.

Redacted. 2007. The CIA Kryptos Sculpture: A summary of previous work and new revelations in working toward its complete solution. https://www.nsa.gov/Portals/70/documents/news-features/declassified-documents/cia-kryptos-sculpture/KRYPTOS_Summary.pdf.

James Reeds. 1977. "Cracking" a random number generator. *Cryptologia*, 1(1):20–26.

Eugene Rogot. 1975a. Cycles for the Gromark Running Key. *The Cryptogram*, 41(1).

Eugene Rogot. 1975b. Gromarks 1-4. *The Cryptogram*, 41(5).

Frank Rubin. 1996. Designing a high-security cipher. *Cryptologia*, 20(3):247–257.

Jim Sanborn. 2009. Oral History interview with Jim Sanborn, 2009 July 14-16. https://www.aaa.si.edu/collections/interviews/oral-history-interview-jim-sanborn-15700.

Klaus Schmeh. 2015. 2015-10-25-Kryptos Workshop. https://youtu.be/25YFYKKKkDo?t=2704.

Christian Schridde. 2020. Kryptos The Cipher Part 4. http://numberworld.blogspot.com/2020/07/kryptos-cipher-part-4.html.

John Schwartz. 2014. Another Kryptos Clue is offered in a 24 year old mystery at the CIA. https://www.nytimes.com/2014/11/21/us/another-kryptos-clue-is-offered-in-a-24-year-old-mystery-at-the-cia.html.

Ferdinando Stehle. 2000. help needed to break KRYPTOS. https://groups.google.com/g/sci.crypt/c/EYd9673EihM.

William Stein. 2007. Sage mathematics software. *http://www.sagemath.org/*.

Jarl Van Eycke. 2015. Schemes 340. http://www.zodiackillersite.com/viewtopic.php?p=38542.