Ciphertext-only attack on $d \times d$ Hill in $O(d13^d)^{\star}$

Shahram Khazaei^a, Siavash Ahmadi^b

^aDepartment of Mathematical Sciences, Sharif University of Technology

^bISSL, Department of Electrical Engineering, Sharif University of Technology

---

Abstract

Hill is a classical cipher which is generally believed to be resistant against ciphertext-only attack. In this paper, by using a divide-and-conquer technique, it is first shown that Hill with $d \times d$ key matrix over $\mathbb{Z}_{26}$ can be broken with computational complexity of $O(d26^d)$, for the English language. This is much less than the only publicly known attack, i.e., the brute-force with complexity of $O(d^326^{d^2})$. Then by using the Chinese Remainder Theorem, it is shown that the computational complexity of the proposed attack can be reduced down to $O(d13^d)$. Using an information-theoretic approach, it is shown that the minimum ciphertext length required for a successful attack increases by a factor of about 7 and 9.8, respectively for these two attacks in comparison with the brute-force attack. This is the only serious attack on Hill since its invention in 1929.

Keywords: Hill cipher, ciphertext-only attack, classical ciphers, Chinese Remainder Theorem, entropy, redundancy

---

1. Introduction

Classical ciphers refer to a type of historically used ciphers which now have fallen into disuse. They are usually divided into substitution and transposition ciphers. In substitution ciphers, groups of letters are systematically replaced throughout the plaintext with other groups of letters. In transposition ciphers, the letters themselves are kept unchanged, but their order within the plaintext is scrambled according to some well-defined scheme. There are several classical ciphers such as Caesar, monoalphabetic and polyalphabetic substitution ciphers, Vigenère square, Great, Morse Code, Pigpen, Scytale, Columnar, Chinese cipher, and so on Sinkov (1966); Denning (1982); Stinson (2005).

Cryptanalysis of classical ciphers is usually a simple work, and most of them are broken not only with Known Plaintext Attacks (KPA), but also with Ciphertext-Only Attacks (COA).

[★]This work has been supported by Iranian National Science Foundation (INSF) under contract no. 92027548 and Sharif Industrial Relation Office (SIRO) under grant no. G931223.

Email addresses: shahram.khazaei@sharif.edu (Shahram Khazaei), s_ahmadi@ee.sharif.edu (Siavash Ahmadi)

There is no need to mention that the existence of some redundancy in the messages (e.g., an English text) is the requisite of breaking a cipher with COA, or otherwise the cipher is information-theoretically secure.

Interestingly, some of these ciphers are still quite resistant against the COA. Our focus on this paper is on the Hill cipher Hill (1929), a polygraphic substitution cipher based on linear algebra, invented by Lester S. Hill in 1929. This cipher uses matrix multiplication of a $d \times d$ secret key matrix with $1 \times d$ plaintext blocks over $\mathbb{Z}_{26}$ to compute the ciphertext blocks. Because of linear property of the cipher, one can easily show that with KPA, the Hill cipher can be simply broken and the matrix of production (i.e., the key matrix) can be extracted. Nevertheless, it is generally accepted that mounting COA on Hill is much harder Wagstaff (2002); Stinson (2005).

The trivial exhaustive key space search on $d \times d$ Hill requires 26^{d^2} matrix multiplications. To the best of our knowledge, no better attack has ever been reported in the literature. We remark that taking into account the invertibility of the key matrix does not lead to a substantially improved attack. It is well-known (e.g., see Overbey et al. (2005)) that the number of invertible $d \times d$ matrices over $\mathbb{Z}_{26}$, i.e., the size of key space $\mathcal{K}$, can be calculated as follows $|\mathcal{K}| = 26^{d^2} \prod_{i=1}^d (1 - 2^{-i})(1 - 13^{-i}) > 0.229 \times 26^{d^2}$. It can be illustrated that this leads to marginal improvement and the asymptotic complexity of the attack does not change, stating that determining the invertibility of a matrix which demands some extra effort has been ignored. Although the complexity of 26^{d^2} for brute-force COA on Hill is common knowledge (see Wikipedia (Accessed Summer 2015)), it seems that there is no report on the possibility of any further improvements in the literature. The security of Hill cipher against COA attack is a common belief repeated in several modern textbooks. Some are content to say that “[Hill] can be difficult to break with a ciphertext-only attack” Stinson (2005), some just say that “A ciphertext-only attack [on Hill] is harder” Wagstaff (2002). In addition the latter states that cryptanalysis based on letter frequency does not work because the Hill cipher encrypts blocks of letters together. Nonetheless, in this paper, we show that cryptanalysis based on letter frequency works just fine. We conclude that the best publicly known COA on Hill cipher requires full search over all $O(26^{d^2})$ possible secret keys. To be more precise, this attack requires $O(26^{d^2})$ operations consisting of $O(d^3)$ multiplications of $d \times d$ square matrices. Therefore, if one does not bother to use fast algorithms for matrix multiplication such as Strassen’s method Strassen (1969), the complexity of trivial brute-force attack will be $O(d^3 26^{d^2})$.

We show that for the English language, a COA can be applied on Hill much faster than full search of key matrix elements (i.e., the brute-force attack). Our first key contribution leads us to a new COA that uses a divide-and-conquer technique by searching over each column of the decryption key matrix, separately. Here, by leveraging the non-uniformity of monograms, we perform the attack. The computational complexity of our COA is $O(d 26^d)$ which is already dramatically lower than the previous ones. In our second contribution, using the Chinese Remainder Theorem (CRT), we devise an attack with computational complexity of $O(d 13^d)$, leveraging non-uniformity of monograms when considered both

modulo 2 and 13. This leads to practical COA on Hill with d as large as say 10 (performing $10 \times 13^{10} \approx 2^{40.4}$ simple operations is affordable on a PC in a tolerable amount of time).

2. Preliminaries

We need the following preliminaries to describe and analyze our attacks, most of which (except Definitions 4 and 5) can be found in Stinson (2005).

English language properties. Various people have estimated frequencies, i.e., probabilities of occurrence of the 26 letters of the English language. Let f_i denote the correct frequency of the i^{th} letter of English alphabet, also known as monogram frequencies. We use the frequencies reported in Stinson (2005) as our reference. If X is a random variable with the same probability distribution for English monograms, then the entropy of X , define as $H_1 = \sum_i f_i \log_2 f_i$, is estimated to be $H_1 \approx 4.1718$. The frequencies of n -gram's and n -gram entropy, denoted by H_n , are similarly defined.

Definition 1 (natural entropy and redundancy). The natural entropy H and the redundancy R of a language with alphabet $\mathcal{P}$ is defined as:

$$H = \lim_{n \rightarrow \infty} \frac{H_n}{n}, \quad (1)$$

$$R = 1 - \frac{H}{\log_2 |\mathcal{P}|}. \quad (2)$$

Experimental results for the natural entropy of English language, with $|\mathcal{P}| = 26$, shows that $1.0 \leq H \leq 1.5$. For later references, we use the more conservative lower bound $H \approx 1.0$.

Definition 2 (unicity distance). The minimum ciphertext length required to break a cipher, i.e., determine the secret key almost uniquely, is called the unicity distance of the cipher.

Using information theory concepts, it is straightforward to show that the unicity distance of a cipher can be calculated according to the following theorem.

Theorem 1. *The unicity distance of a cipher with key space $\mathcal{K}$, over a plaintext space with alphabet $\mathcal{P}$ and redundancy R , is:*

$$n_0 \approx \frac{\log_2 |\mathcal{K}|}{R \log_2 |\mathcal{P}|}.$$

We remark that many cryptanalysis techniques mount a COA only using the non-uniformity of monograms, instead of taking advantage of the full redundancy in the plaintext. The cryptanalysis techniques based on monogram frequencies usually use index of coincidence, defined below, to measure how well a string of English alphabet matches the English language in terms of monograms.

Definition 3 (index of coincidence). The index of coincidence (IC) of a string P over English alphabet with the observed (normalized) frequency $\hat{f}_i$ for the i^{th} letter (i.e., $\sum_i \hat{f}_i = 1$) is defined as:

$$\text{IC}(P) = \sum_i \hat{f}_i f_i$$

For many ciphers, when a given ciphertext is decrypted using a random wrong key, it is reasonable to assume that the decrypted string is uniformly random. We call this assumption Simple Uniform Wrong Key Decryption (SUWKD). Under SUWKD assumption, using statistical hypothesis testing similar to those used in correlation attacks Siegenthaler (1984), one can argue that the optimum criteria to measure how well a string matches the English language in terms of monograms, is the index of maximum likelihood, defined as follows.

Definition 4 (index of maximum likelihood). The index of maximum likelihood (IML) of a string P over English alphabet with the observed (normalized) frequency $\hat{f}_i$ for the i^{th} letter is defined as:

$$\text{IML}(P) = - \sum_i \hat{f}_i \log_2 f_i$$

To the best of our knowledge, this is the first time that IML is used for cryptanalysis of classical ciphers instead of IC. We formalize the notion of how well a string over English alphabet matches the English language using the following definition. We ignore to provide a similar definition based on IC due to its non-optimality.

Definition 5 (monogram-wise meaningful string). We say that a string P over English alphabet is monogram-wise meaningful if the corresponding $\text{IML}(P)$ is greater than a certain threshold.

Let us discuss how a ciphertext C of a cipher with key space $\mathcal{K}$ can be broken using brute-force attack and based on monogram frequencies. To find a few candidates as the potential correct keys, one approach is to decrypt the ciphertext with all possible candidate keys $K \in \mathcal{K}$ and report those whose corresponding decrypted string is monogram-wise meaningful for an appropriately chosen threshold. Another approach is to report the key (or a few keys) with the highest IML for the corresponding decrypted string. Essentially, if the threshold in Definition 5 is chosen properly based on the ciphertext length and the key space size, these two approaches are statistically the same. The interested reader is referred to Siegenthaler (1984) for justification.

3. Description of Hill

Without loss of generality, we assume that for Hill cipher, the plaintext space is the set of all meaningful English strings whose length is a multiple of an integer d . Each character is naturally interpreted as an element of $\mathbb{Z}_{26}$. To encrypt a plaintext $P = (p_1, p_2, \dots, p_{md})$

using a $d \times d$ key matrix K over $\mathbb{Z}_{26}$, it is first divided into m blocks of d characters. Let $P_i = (p_{(i-1)d+1}, p_{(i-1)d+2}, \dots, p_{(i-1)d+d})$ denote the i^{th} block of the plaintext for $i = 1, 2, \dots, m$. The corresponding ciphertext block C_i is then calculated as $C_i = P_i K$ to construct the final ciphertext $C = (C_1, C_2, \dots, C_m)$.

COA on Hill. Although Hill is easily broken with KPA, no attack has been reported faster than brute-force attack in COA model, with the assumption that the plaintext is an English text. The existing redundancy of English text can be used to mount COA on Hill, e.g, using a brute-force attack. In this case, we have $|\mathcal{K}| \approx 26^{d^2}$, $|\mathcal{P}| = 26$, and $H \approx 1.0$. Therefore, $R \approx 0.787$ and by Theorem 1, the attack can determine the secret key almost uniquely if the ciphertext is of length at least:

$$n_0 \approx \frac{\log_2 26^{d^2}}{0.787 \log_2 26} \approx 1.27d^2 \quad (3)$$

In the rest of the paper, we show how our new attacks can be applied on Hill cipher in the COA model much faster than brute-force attack, with the required ciphertext length slightly increased.

4. COA on Hill using monograms

In this section, we describe a brute-force attack on Hill using monograms only (instead of the whole language redundancy). We then further improve the attack using the divide-and-conquer technique. However, this will not result in finding the correct secret decryption matrix and, instead, we will find the secret matrix uniquely up to an unknown permutation of its columns. The correct order of the columns can then be determined using bigram frequencies very efficiently which is equivalent to breaking a permutation cipher. First we define the notion of a representative key. Before definition, note that a ciphertext block C is decrypted under a decryption key matrix K^{-1} according to $P = CK^{-1}$. We abuse the notation and similarly denote the decryption of a ciphertext C of an arbitrary length, but a multiple of d , by $P = CK^{-1}$.

Definition 6 (representative key). For a given ciphertext C , we say that a (candidate) decryption key matrix K^{-1} is a representative key if the decrypted string $P = CK^{-1}$ is monogram-wise meaningful (for a certain threshold).

4.1. brute-force attack on Hill

In order to characterize the representative keys for the Hill cipher, we need the following theorem, whose proof is straightforward based on the given definitions so far.

Theorem 2 (permutation rule). *For a given ciphertext C , if a matrix K^{-1} is a representative key, then so is the matrix produced by any arbitrary permutation of the columns of K^{-1} .*

Consequently, for a long enough ciphertext and for a well-chosen threshold, all the $d!$ matrices derived from the correct key matrix are the sole representative keys. We conclude that in order to find all the representative keys, one can exhaust all the $O(26^{d^2}/d!)$ matrices — which are equivalent in terms of column permutation — instead of a brute-force attack over all possible matrices. Let us discuss the unicity distance of the attack, i.e., the minimum ciphertext length to determine the $d!$ correct representative keys almost uniquely. We need to plug $|\mathcal{K}| \approx 26^{d^2}/d!$, $|\mathcal{P}| = 26$ and $R \approx 0.1107$ (corresponding to $H \approx 4.1758$) in Theorem 1 to get $\frac{\log_2(26^{d^2}/d!)}{0.1107 \log_2 26} \approx 8.96d^2 - O(\log d)$. In other words, if the ciphertext length is about this amount, then the secret matrix is almost uniquely determined up to an unknown permutation over its columns. The bigram frequencies can then be used to determine the correct order of the columns of a representative key with computational complexity of $O(d^2)$; the complexity of which is ignorable in comparison with the overall exponential complexity of the attack. Therefore, we can ignore the complexity of the permutation attack and consider finding a correct representative key as the end goal. To summarize, the computational complexity of the attack, without bothering oneself to use fast matrix multiplication algorithms, is $O(d^3 26^{d^2}/d!) = O(d^2 26^{d^2}/\log d)$.

4.2. A divide-and-conquer attack on Hill

We describe how a divide-and-conquer technique can be used to transform the brute-force attack into a new attack with remarkable improvement in computational complexity. This attack is based on three key observations:

- i) Let K_j^{-1} denote the j^{th} column of a candidate matrix K^{-1} . Given a ciphertext block C_i , the j^{th} element of the decrypted string block $P_i = C_i K^{-1}$, which we denote by $p_i^{(j)}$, can be calculated according to $p_i^{(j)} = C_i K_j^{-1}$.
- ii) We expect that the monogram frequencies are still observed if we decimate an English language string (e.g., by keeping the letters which are d positions apart). In other words, if the number of ciphertext blocks C_i 's is sufficient, for a column K_j^{-1} of the correct decryption matrix K^{-1} , the sequence of $p_i^{(j)} = C_i K_j^{-1}$'s is monogram-wise meaningful.
- iii) We conclude that each column of decryption key matrix can be found separately and independently by trying all $(2^d - 1)(13^d - 1) \approx 26^d$ possibilities for a column. Since there is no superiority in considering each column, guessing a single column of K^{-1} actually reveals all the correct columns, provided that the ciphertext is long enough.

Using Theorem 1, the number of decrypted letters sufficient to almost uniquely determine a column, in fact all the d columns, of the decryption matrix can be calculated as $\frac{\log_2 26^d}{0.1107 \log_2 26} \approx 8.96d$. Notice, however, that to obtain the above amount of decrypted letters, one needs a ciphertext of length $n_0^{(26)} \approx 8.96d^2$.

Compared with the trivial brute-force attack, which takes advantage of the full redundancy of the English plaintext, the required minimum length to mount a successful attack increases by a factor of $n_0^{(26)}/n_0 \approx 7$; see Eq. (3). Now, let us discuss the complexity of the attack. Our divide-and-conquer attack performs $O(26^d)$ vector by matrix multiplications. Therefore, a naive implementation of the attack has a computational complexity of $O(d^2 26^d)$. By eliminating repeated calculations, an improved attack with complexity of $O(d 26^d)$ is presented in Algorithm 1.

Algorithm 1 Divide-and-conquer attack on Hill in $O(d 26^d)$

Require: A ciphertext C of length $n = md$

Ensure: A representative decryption matrix K^{-1}

- 1: Divide C into m blocks $C_1, \dots, C_m$
  - 2: **for** $t = 0$ **to** $d - 1$ **do**
  - 3:   **for** $i = 1$ **to** m **do**
  - 4:   Set $d_{i,t}$ to the modulo 26 sum of the first $t + 1$ elements of C_i
  - 5: Set K^{-1} to an arbitrary $d \times d$ matrix
  - 6: Set I to an all $-\infty$ vector of length d
      {the j 'th element of I corresponds to the IML of the j 'th column of K^{-1} }
  - 7: **for** $i = 1$ **to** m **do**
  - 8:   Set $p_i = 0$
      { p_i is a decrypted letter of C_i using an all-zero guess for a column of K^{-1} }
  - 9: **for** all $d \times 1$ vectors x (except the all-zero vector) in lexicographical order **do**
  - 10:   Let t denote the maximum number of zeros at the top of x .
  - 11:   **for** $i = 1$ **to** m **do**
  - 12: $p_i = p_i + d_{i,t} \bmod 26$
  - 13: $iml_x = \text{IML}(p_1, \dots, p_m)$
  - 14:   Let y be a column of K^{-1} with the smallest index $iml_y = \min I$
  - 15:   **if** $iml_y < iml_x$ **and** x is not all-zero modulo 2 or 13 **then**
  - 16:   Replace the column y in K^{-1} with x
  - 17:   Replace iml_y in I with iml_x
-

5. A CRT based divide-and-conquer attack on Hill

Based on the following Observation, the computational complexity of the divide-and-conquer attack can be improved using the Chinese Remainder Theorem (CRT).

Observation 1 ($\mathbb{Z}_{13}$ and $\mathbb{Z}_2$ entropy). Let X denote a random variable over $\mathbb{Z}_{26}$ with the same probability distribution for English letters. The $\mathbb{Z}_{13}$ entropy of English monograms is defined to be the entropy of the random variable $(X \bmod 13)$. Similarly, we define the $\mathbb{Z}_2$ entropy. A simple calculation shows that $\mathbb{Z}_{13}$ and $\mathbb{Z}_2$ entropy of English monograms

are 3.4052 and 0.9865, respectively. The corresponding redundancies are then 0.0798 and 0.0135.

With this observation, the same procedure for divide-and-conquer attack can be done in order to find the columns of the decryption key matrix modulo 2 and 13. To be more precise, for example, in order to find the decryption key matrix columns modulo 13, the ciphertext is first reduced modulo 13. Then all possible $13^d - 1$ values for a column modulo 13 are tried using the brute-force attack. Finally, the best d candidates for the probable columns modulo 13 are identified. The minimum ciphertext length required to almost uniquely find a representative decryption key matrix modulo 13 can be approximated as

$$n_0^{(13)} \approx \frac{\log_2 13^d}{0.0798 \log_2 13} \times d \approx 12.5d^2.$$

A similar procedure can be performed to find a representative decryption key matrix modulo 2, requiring a ciphertext length of length $n_0^{(2)} \approx \frac{\log_2 2^d}{0.0135 \log_2 2} \times d \approx 74d^2$.

Interestingly, to find a representative key modulo 26, the attack can be devised in two different ways using the CRT. Although the computational complexities are the same, the required ciphertext lengths for a successful attack are different.

Combinational attack. Perform the divide-and-conquer attack modulo 13 and 2, and attain a representative decryption key matrix over $\mathbb{Z}_{13}$ and one over $\mathbb{Z}_2$. For performing these attacks using Theorem 1, one needs a ciphertext of length $12.5d^2$ for $\mathbb{Z}_{13}$ and $74d^2$ for $\mathbb{Z}_2$. Now, combine each column j of the representative key over $\mathbb{Z}_{13}$ with all the d columns of the representative key over $\mathbb{Z}_2$ and extract d vectors for the j 'th column over $\mathbb{Z}_{26}$ using the CRT. Then by calculating IML for all of these d possibilities, the vector with the largest index is considered as the j 'th column of the representative key over $\mathbb{Z}_{26}$. The computational complexity of this method is $O(d13^d + d2^d + d^4) = O(d13^d)$, and it needs a ciphertext of length at least $74d^2$.

Lifting attack. Perform the divide-and-conquer attack over modulo 13 only, and attain a representative key over $\mathbb{Z}_{13}$. Then for each column j of the matrix do the following steps to lift the $\mathbb{Z}_{13}$ representative key into one over $\mathbb{Z}_{26}$:

- a) combine all the $2^d - 1$ non-zero vectors over $\mathbb{Z}_2$ with the j 'th column and compute the corresponding vectors over $\mathbb{Z}_{26}$ using the CRT.
- b) calculate the IML for each one and choose the vector with the largest index as the j 'th column of the representative key matrix over $\mathbb{Z}_{26}$.

Now a representative key over $\mathbb{Z}_{26}$ is attained. The computational complexity of this method is $O(d13^d + d^3 2^d) = O(d13^d)$, but it requires a ciphertext of length at least $12.5d^2$, i.e., only a factor of $n_0^{(13)}/n_0 \approx 9.8$ larger than the trivial brute-force attack.

6. Conclusion

Hill is a classical encryption system which is generally conceived as a ciphertext-only attack resistant block cipher. In this paper we presented a new COA on Hill using divide and conquer technique and the redundancy of the monograms only. Our attack then turns the ciphertext into the output of a permutation cipher, which can be easily broken using the redundancy of bigrams. The computational complexity of the attack is $O(d26^d)$ which is dramatically lower than the computational complexity of brute-force attack on Hill. Also, by using Chinese Remainder Theorem, we improved the divide-and-conquer COA on Hill to the computational complexity of $O(d13^d)$ with the cost of a slightly more data complexity. Beating our results or proving its optimality based on some reasonable computational complexity assumptions remains an open problem.

Acknowledgment. The initial idea of the basic divide-and-conquer attack was proposed by the following student in an introductory course in cryptography at the Department of Mathematical Sciences at Sharif University of Technology offered by the first author: Navid Alamati, Mohammad-Esmaeil Hasani, and Morteza Hasanvand.

Denning, D. E. R., 1982. Cryptography and data security. Addison-Wesley Longman Publishing Co., Inc.

Hill, L. S., 1929. Cryptography in an algebraic alphabet. In: American Mathematical Monthly. pp. 306–312.

Overbey, J., Traves, W., Wojdylo, J., 2005. On the keyspace of the hill cipher. Cryptologia 29 (1), 59–72.
URL <http://dx.doi.org/10.1080/0161-110591893771>

Siegenthaler, T., 1984. Correlation-immunity of nonlinear combining functions for cryptographic applications. IEEE Transactions on Information Theory 30 (5), 776–780.
URL <http://dx.doi.org/10.1109/TIT.1984.1056949>

Sinkov, A., 1966. Elementary cryptanalysis: a mathematical approach. Vol. 22 of New mathematical library. With a supplement by Paul L. Irwin. Reissued in 1975 and 1980.

Stinson, D. R., 2005. Cryptography: theory and practice. CRC press.

Strassen, V., 1969. Gaussian elimination is not optimal. Numerische Mathematik 13 (4), 354–356.

Wagstaff, Samuel S., J., 2002. Cryptanalysis of number theoretic ciphers. CRC Press.

Wikipedia, E., Accessed Summer 2015. Hill Cipher. https://en.wikipedia.org/wiki/Hill_cipher.