Katelyn Payton-Page, Chad Etheredge, Lee Beadle, Dillon Borden

# Cryptography Project Report

May 9, 2017

## Part I: Cryptanalysis Project 2

The Hill Cipher: a classical cipher utilizing linear algebra to encrypt messages. It was created by Lester S. Hill in 1929 []. The advantage of the Hill Cipher, over other classical ciphers, is that it is polygraphic. This means that the plaintext can be substituted and encrypted in groups. In other classical cyphers, letters are substituted individually instead of in groups. Being polygraphic makes the Hill cypher less susceptible to frequency analysis, as it makes the frequency distribution flatter than those given by mono-alphabetic ciphers.

To encrypt messages using the Hill Cipher, first choose an invertible $n \times n$ key matrix. A matrix is invertible in modulo 26 if it has a determinant that is not zero and an inverse determinant that also exists in modulo 26. This key matrix is then multiplied by subsets of $n \times 1$ plaintext vectors to get a one-to-one correspondence of plaintext to ciphertext. In order for matrix multiplication to work for this cryptosystem is the number of columns in the key matrix must equal the number of rows in the plaintext vectors. If the number of plaintext characters $mod 26$ is not zero, then pad plaintext with Z. The formula for encryption is the following

$$K_{n\times n} \cdot \vec{P}_{n\times 1} = \vec{C}_{n\times 1}.$$

Where $K$ is the key matrix, $\vec{P}$ is the plaintext vector, and $\vec{C}$ is the ciphertext vector.

Now to decrypt a message, the inverse of the key matrix, $K$, needs to be found. Then take that inverse matrix and multiply it by the ciphertext, which is broken into a $n \times 1$ component vectors to get back to the original plaintext vectors. If there are extra z's at the end, these are paddings and should be removed. The formula for decrypting is the following:

$$(K_{n\times n})^{-1} \cdot \vec{C}_{n\times 1} = \vec{P}_{n\times 1}.$$

The following examples are specific to the cryptosystem that was given to analyze. It follows that the Hill Cipher with a small variation, after matrix multiplication, will also be adding an offset vector, $\vec{V}$.

Encryption Example:

Given a key matrix $K$, an offset vector $\vec{V}$, and plaintext $P$. Find the ciphertext, $C$.

$$K_{3\times 3} = \begin{bmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{bmatrix}$$

$$\vec{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$P$ = LIONS

Since $n = 3$ and LIONS is not a multiple of three, add a Z to the end of LIONS, so

$P$ = LIONSZ

Making the plaintext into a vector:

$$\vec{p_1} = \begin{bmatrix} L \\ I \\ O \end{bmatrix} = \begin{bmatrix} 11 \\ 8 \\ 14 \end{bmatrix}$$

$$\vec{p_2} = \begin{bmatrix} N \\ S \\ Z \end{bmatrix} = \begin{bmatrix} 13 \\ 18 \\ 25 \end{bmatrix}$$

Now determine if $K$ is invertible $mod 26$.

$$det(K) = 441 \ mod \ 26 = 25$$
$$25^{-1} \ mod \ 26 = 25$$

So $K$ is invertible in $mod\ 26$.
So the plaintext vectors can then be encrypted.

$$K \cdot \vec{p_1} + \vec{V} \ mod \ 26 = \begin{bmatrix} 9 \\ 7 \\ 16 \end{bmatrix} = \begin{bmatrix} J \\ H \\ Q \end{bmatrix}$$

$$K \cdot \vec{p_2} + \vec{V} \ mod \ 26 = \begin{bmatrix} 7 \\ 12 \\ 23 \end{bmatrix} = \begin{bmatrix} H \\ M \\ X \end{bmatrix}$$

Therefore, $C = JHQHMX$.

Decryption Example:

Given an invertible key matrix $K$, offset $\vec{V}$ and ciphertext $C$, find the plaintext $P$

$$K_{3\times3} = \begin{bmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{bmatrix}$$

$$\vec{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C = JHQHMX$$

$$\vec{c_1} = \begin{bmatrix} J \\ H \\ Q \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \\ 16 \end{bmatrix}$$

$$\vec{c_2} = \begin{bmatrix} H \\ M \\ X \end{bmatrix} = \begin{bmatrix} 7 \\ 12 \\ 23 \end{bmatrix}$$

Invert the matrix $K \; mod \; 26$

$$K^{-1} = \begin{bmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{bmatrix}$$

When the source code is analyzed, the ciphertext vectors are transposed to column vectors. It is then multiplied by $K$ on the opposite side of the encryption. So in order for this to be conformed to the rules of matrix multiplication, the inverted key matrix must also be transposed.

$$(K^{-1})^T \; mod \; 26 = \begin{bmatrix} 8 & 21 & 21 \\ 5 & 8 & 12 \\ 10 & 21 & 8 \end{bmatrix}$$

Now all of the information that is needed for decryption is known.

So now remove the offset vector from the ciphertext vectors.

$$\vec{c_1} - \vec{V} = \begin{bmatrix} 8 \\ 6 \\ 15 \end{bmatrix}$$

$$\vec{c_2} - \vec{V} = \begin{bmatrix} 6 \\ 11 \\ 22 \end{bmatrix}$$

Now decrypt the ciphertext.

$$(\vec{c_1})^T \; (K^{-1})^T = \begin{bmatrix} 11 \\ 8 \\ 14 \end{bmatrix} = \begin{bmatrix} L \\ I \\ O \end{bmatrix}$$

$$(\vec{c_2})^T \ (K^{-1})^T = \begin{bmatrix} 13 \\ 18 \\ 25 \end{bmatrix} = \begin{bmatrix} N \\ S \\ Z \end{bmatrix}$$

Therefore, the plaintext is LIONSZ. Which is the desired plaintext.

## Part II: Known Plaintext Cryptanalysis of the 3x3 Extended Hill Cipher

The following is a method of generating an unknown matrix key and unknown offset vector from an Extended Hill Cipher cryptosystem. The method utilizes a known plaintext/ciphertext attack to create a system of linear equations solved with Matrix Reduction operations. The Hill Cipher is particularly vulnerable to known plaintext attacks due to its linear structure. Despite the fact that there is an added offset to introduce a non-linear component to the cryptosystem, the system is still vulnerable to linear algebra attack methods.

To begin the cryptanalysis, identify the letters from the plaintext with their corresponding letters from the ciphertext. Then split the plaintext and ciphertext into groups of three. Each plaintext tri-graph is represented as $\vec{P}$, which contains their three letters as their corresponding number values. Now each ciphertext tri-graph is represented as $\vec{C}$, which also contains the ciphertext's three letters as their corresponding number values. Let $K$ be represented as the key matrix.

$$K = \begin{bmatrix} k_1 & k_4 & k_7 \\ k_2 & k_5 & k_8 \\ k_3 & k_6 & k_9 \end{bmatrix}$$

Let the unknown offset values be $\vec{V}$, which is composed of three elements.

$$\vec{V} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Now that everything is labeled, there is an equation created such that $\vec{P} \cdot K + \vec{V} = \vec{C}$. So the equation will look like,

$$\vec{P} \cdot \begin{bmatrix} k_1 & k_4 & k_7 \\ k_2 & k_5 & k_8 \\ k_3 & k_6 & k_9 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \vec{C}$$

Each of the tri-graphs from the plaintext and their corresponding ciphertext can be observed as three equations. So for example, $\vec{P}$'s first tri-graph is the letters $A\ D\ I$ (where $A = 0$, $D = 3$, and $I = 8$) and is placed into the following three equations,

$$\begin{gathered} 0 \cdot k_1 + 3 \cdot k_2 + 8 \cdot k_3 + v_1 = 3 \\ 0 \cdot k_4 + 3 \cdot k_5 + 8 \cdot k_6 + v_2 = 18 \\ 0 \cdot k_7 + 3 \cdot k_8 + 8 \cdot k_9 + v_3 = 17. \end{gathered}$$

So in the $3 \times 3$ key matrix, $K$, there is nine unknown values. Additionally, due to the offset $\vec{V}$, there is three more unknown variables. So there is a total number of twelve unknown variables. Now to generate $K$ and $\vec{V}$, the twelve unknown variables need to be solved. Therefore, choose four tri-graphs to represent the system of twelve linear equations and convert each letter to their specific number values (A = 0, B = 1, C = 2, ...). For each tri-graph there will be nine unknown $K$ variables whose coefficients will be the specific number of the letters in the Plaintext. Also, there will be the three unknown $\vec{V}$ variables, whose coefficients will be one. Lastly, all this will equal the specific number values of the letters from the Ciphertext. So we will have the following four tri-graphs,

$$
\begin{gathered}
ADI \to DSR \\
0 \cdot k_1 + 3 \cdot k_2 + 8 \cdot k_3 + v_1 = 3 \\
0 \cdot k_4 + 3 \cdot k_5 + 8 \cdot k_6 + v_2 = 18 \\
0 \cdot k_7 + 3 \cdot k_8 + 8 \cdot k_9 + v_3 = 17,
\end{gathered}
$$

$$
\begin{gathered}
SPL \to MSI \\
18 \cdot k_1 + 15 \cdot k_2 + 11 \cdot k_3 + v_1 = 12 \\
18 \cdot k_4 + 15 \cdot k_5 + 11 \cdot k_6 + v_2 = 18 \\
18 \cdot k_7 + 15 \cdot k_8 + 11 \cdot k_9 + v_3 = 8,
\end{gathered}
$$

$$
\begin{gathered}
AYE \to OPL \\
0 \cdot k_1 + 24 \cdot k_2 + 4 \cdot k_3 + v_1 = 14 \\
0 \cdot k_4 + 24 \cdot k_5 + 4 \cdot k_6 + v_2 = 15 \\
0 \cdot k_7 + 24 \cdot k_8 + 4 \cdot k_9 + v_3 = 11,
\end{gathered}
$$

$$
\begin{gathered}
DEQ \to XLJ \\
3 \cdot k_1 + 4 \cdot k_2 + 16 \cdot k_3 + v_1 = 23 \\
3 \cdot k_4 + 4 \cdot k_5 + 16 \cdot k_6 + v_2 = 11 \\
3 \cdot k_7 + 4 \cdot k_8 + 16 \cdot k_9 + v_3 = 9.
\end{gathered}
$$

Next, construct nine more equations from three different tri-graphs. This system of linear equations can be inserted into a $12 \times 13$ matrix that looks like this,

$$
A = \begin{bmatrix}
0 & 3 & 8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 3 \\
0 & 0 & 0 & 0 & 3 & 8 & 0 & 0 & 0 & 0 & 1 & 0 & 18 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 8 & 0 & 0 & 1 & 17 \\
18 & 15 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 12 \\
0 & 0 & 0 & 18 & 15 & 11 & 0 & 0 & 0 & 0 & 1 & 0 & 18 \\
0 & 0 & 0 & 0 & 0 & 0 & 18 & 15 & 11 & 0 & 0 & 1 & 8 \\
0 & 24 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 14 \\
0 & 0 & 0 & 0 & 24 & 4 & 0 & 0 & 0 & 0 & 1 & 0 & 15 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 24 & 4 & 0 & 0 & 1 & 11 \\
3 & 4 & 16 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 23 \\
0 & 0 & 0 & 3 & 4 & 16 & 0 & 0 & 0 & 0 & 1 & 0 & 11 \\
0 & 0 & 0 & 0 & 0 & 0 & 3 & 4 & 16 & 0 & 0 & 1 & 9
\end{bmatrix}
$$

This matrix is the augmented matrix $A$. Groups of three rows represents one tri-graph of the plaintext to the ciphertext. Columns one through nine denotes the corresponding coefficients of the key values, ($k_1$, $k_2$, $k_3$, $k_4$, $k_5$, $k_6$, $k_7$, $k_8$, and $k_9$). The columns ten through twelve, denote the coefficients of $\vec{V}$ values, ($v_1$, $v_2$, and $v_3$). Then the last column, column thirteen, is the results. The last column is produced by the ciphertext values, which is located on the right-side of the equations.

Once the $12 \times 13$ matrix is constructed, perform row operations to reduce the matrix to Reduced Echelon Form. In this analysis, Sage was used to reduce the given matrix. By Reduced Echelon Form, if a matrix was to have all zeros in a column, then the tri-graph that includes the equation fails. Then that tri-graph needs to be replaced with another tri-graph since it is linearly dependent. If a row contains all zeros, then one or more of the equation variables share linear dependency. In this case, a different combination of tri-graph equations can be tested. Therefore, it is beneficial to have $3^2$ groups of 3 tri-graphs. So in general, it is better to have $n^2$ groups of $n$-graphs, where $n^2$ is the size of $K$.

$$rref(A) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-565}{921} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{313}{307} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{799}{307} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{352}{921} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-101}{307} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-300}{307} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \frac{-34}{921} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \frac{-142}{307} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \frac{-285}{307} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \frac{-6410}{307} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \frac{8229}{307} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \frac{7925}{307} \end{bmatrix}$$

If the matrix, $A$, is properly reduced, then the results column (column thirteen) will list the $K$ values ($k_1$ - $k_9$) and the $\vec{V}$ values ($v_1$, $v_2$, and $v_3$). Compute their representations by modulo 26. Looking over the results column, every three integers will form a column in $K$ and the last three values are elements in $\vec{V}$.

$$K = \begin{bmatrix} 3 & 6 & 4 \\ 5 & 15 & 18 \\ 17 & 8 & 5 \end{bmatrix} \vec{V} = \begin{bmatrix} 8 \\ 13 \\ 1 \end{bmatrix}$$

Once $K$ and $\vec{V}$ have been computed, the values can then be checked through the initial equations from the tri-graphs $mod\ 26$. Since, the values of the ciphertext are known, then it is possible to compare the results. Likewise, the same enciphering/deciphering system can be solved with $K$ and $\vec{V}$. In this case, the $K$ values will not work and the transpose of $K$ may be necessary before $K$ is discarded for another possible solution.

## Part III: Possible Adaptation to Ciphertext Only Attacks

For the given crytosystem, a $3 \times 3$ key matrix, $K$, and a vector, $\vec{V}$, are necessary for encryption and decryption. Despite the addition of the vector, $\vec{V}$, the cryptosystem still falls prey to a known plaintext attack. If, however, the attacker only had access to the ciphertext, attacking the system to determine the key matrix, $K$, would become much more complex. In a known plaintext attack, there are twelve unknown variables. In a ciphertext only attack, there are twenty-one unknown variables. Frequency analysis of trigrams can be used to help reduce the number of unknown variables. However, this approach requires some degree of guesswork and assumption. First, there must be a large body of ciphertext to analyze for the frequency of trigrams. Second, the highest frequency ciphertext trigrams must be assumed to be the highest occuring trigrams of the English language. By making this assumption, the number of unknown variables is reduced from twenty-one to twelve unknown variables. A system of equations can be developed from the addition of the possible plaintext values. A $12x13$ matrix can be constructed and reduced to solve. The results column represents a possible key for the matrix and possible vector values. Since the highest occuring ciphertext trigrams were assumed to be the highest english language trigrams, the solution is not guaranteed to be the correct key. If, after checking and the key is incorrect, a trigram's equations can be matched to another possible plaintext/ciphertext correspondence. This method can be repeated, or iterated through, until ciphertext decryption results in a possible plaintext translation. However, it should be noted that due to the polygraphic nature inherit to the Hill Cipher, frequency analysis is not a strong approach. This is due to the possibility of trigrams not occuring on multiples of three or multiples of the key size. Ultimately, this will skew frequency analysis. Also, with the addition of another unknown variable of vector $V$, the problem is further complicated.