Cryptography: Matrices and Encryption

By: Joseph Pugliano and Brandon Sehestedt

Abstract

The focus of this project is investigating how to generate keys in order to encrypt words using Hill Cyphers. Other forms of encryption will also be looked at, such as the Enigma encryption from World War II, and comparisons are drawn between the two. The effectiveness of the encryptions is studied, and alternate, more secure means of encryption are presented as well.

Introduction

Lester Hill invented the Hill Cypher in 1929. At the time, the cypher was one of the first to be able to operate on more than three letters or symbols during a single encryption or decryption operation. However, as the number of symbols increases, the arithmetic required to perform the encryption becomes more and more difficult, as we will see shown later on.

Key Matrices

The most vital component of the Hill Cypher is the key matrix. The key matrix is used to encrypt the messages, and its inverse is used to decrypt the encoded messages. It is important that the key matrix be kept secret between the message senders and intended recipients. If the key matrix or its inverse is discovered, then all intercepted messages can be easily decoded.

However, not just any matrix can be used as a key. First, the matrix must be invertible, which means its determinant cannot be zero. Additionally, the

determinant of the key matrix must not have any common factors with the modular base used in the encryption and decryption. The modular base corresponds to the number of symbols being used in the encryption. For example, if we use the letters A-Z as our available symbols, then the modular base of the encryption is 26. This means that the determinant of the key matrix cannot be divisible by 2 or 13. A common solution is to add symbols to the available alphabet, such as '!' or '?' or any number of other symbols, in order to make the modulus prime.

The key matrix must have the characteristics outlined above, or the cypher will be unusable. Technically, a message can be encrypted using any square matrix. However, unless it has the characteristics of a key matrix, the encoded message will be impossible to decrypt, causing the original message to be lost.

Encryption Example

Let us demonstrate a simple encryption. For ease, we will use a 2x2 key matrix for the encryption. The message we want to encrypt is "UTES". Since our key matrix is 2x2, we must split the message into chunks of two letters. This is true for all matrices, so for an *N*x*N* matrix, the message would be split into chunks of *N* letters.

Each letter in the message is assigned a numerical value, ranging from 0 up to the modular base. For our encryption, we will use the letters A-Z (so the modulus is 26). While in practice the letters can be assigned any value (as long as the sender and recipient both know the corresponding values), we will use the straightforward approach. So, A = 0, B = 1, C = 2, etc. all the way to Z = 25. Below we see how our message is assigned:

$$UTES = 20\ 19\ 4\ 18 = \begin{pmatrix}20\\19\end{pmatrix}, \begin{pmatrix}4\\18\end{pmatrix}$$

Our key matrix K for this example will be:

$$K = \begin{pmatrix}4 & 1\\3 & 7\end{pmatrix}$$

Now for each chunk of our message, we perform a matrix multiplication against the key matrix. As we will see, the resulting values will fall far outside the range of values we assigned to each letter. To remedy this, we will take each of the resulting values modulo 26:

$$\begin{pmatrix}4 & 1\\3 & 7\end{pmatrix}\begin{pmatrix}20\\19\end{pmatrix} = \begin{pmatrix}99\\193\end{pmatrix} Mod\ 26 = \begin{pmatrix}21\\11\end{pmatrix} = \begin{pmatrix}V\\L\end{pmatrix}$$

$$\begin{pmatrix}4 & 1\\3 & 7\end{pmatrix}\begin{pmatrix}4\\18\end{pmatrix} = \begin{pmatrix}34\\138\end{pmatrix} Mod\ 26 = \begin{pmatrix}8\\8\end{pmatrix} = \begin{pmatrix}I\\I\end{pmatrix}$$

We have now successfully encrypted our original message. Starting with the message "UTES", after multiplying against our key matrix our encrypted message is "VLII". While we will not discuss in depth the decryption process, it is very similar to the encryption, where the inverse of the key matrix is multiplied against the encoded message.

This example was very simple as we used a 2x2 matrix, but it should be easy to see that this process can quickly become quite complex as the dimension of the key matrix increases. For example, using a 10x10 key matrix would require a considerable amount of work, compared to the 2x2 example provided. For this reason, as the dimension of the key matrix increases, the more secure the Hill Cypher becomes. For perspective, for an *N*x*N* matrix using our encryption example, there are 26^{N^2} possibilities. So, for a 2x2 matrix, that provides over 456,000 unique matrices, while a simple increase to a 3x3 matrix can provide of 5 trillion unique matrices. While not all of those are qualified to be key matrices, we see that the possibilities are quite large.

Cracking the Cypher

As discussed above the Hill Cypher does a great job of encrypting a message. When it first came to be, cracking the cypher was not an easy task. With so many possible key matrices and the fact that the cypher covers up repeated letters cracking the message by hand took a lot of time. An attack that was developed to

crack the Hill Cypher is known as the “Plain Text Attack”. The Plain Text Attack is when the attacker has access to both the actual message that was sent as well as the encrypted message. The attacker can then encode the message again and again until he finds the key that gives you the same encrypted message that you intercepted. Once you have the key you can decrypt the rest of the messages that use that same key.

The Plain Text Attack is not the only way to crack the Hill Cypher. With the advancement of computers over time the Hill Cypher became vulnerable to computer hacking. A computer can test out thousands of keys in a matter of seconds which makes it possible to find the correct key in a matter of minutes. Although the Hill Cypher is no longer used for encryptions because of its vulnerability it was a building block for several other encryptions.

Alternative: Enigma

The Enigma machine was famously used during World War II by the Germans to send and receive coded messages. The machine originally was a lot like a Hill Cypher. It had three rotors that would spin and assign a letter to a different letter when pressed. This design of the machine was cracked by a group of Polish mathematicians by using Linear Algebra. When the war broke out the Germans created a more effective enigma machine that included wires in the front to mix the letters up even more and make the code that much more difficult to crack. With the new features the enigma machine had 15 x 10^18 different combinations. This was far too many for any person to solve by hand in a matter of days.

The way that enigma was cracked was by using a computer and the Plain Text Attack that we talked about earlier. The British new a certain word that would be in the first message sent out every day and then could crack the rest of the message. Without the plain text attack a modern computer could still take up to a year to crack an enigma message.

Alternative: RSA

As we have shown, especially with the computing power available at our fingertips today, the Hill Cypher is not a very secure form of encryption. As an alternative form of encryption, we will briefly look at RSA, which is widely used and accepted as a very effective and secure encryption.

RSA employs the use of extremely large prime numbers as keys. Today these keys are usually 2048 bits long, which would correspond to a decimal number that is 617 digits. RSA relies on the fact that there is no known algorithm to efficiently factor such large numbers.

Since RSA does not directly use linear algebra or matrices, the inner workings of the encryption will not be discussed. Simply put, each sender and recipient has both a public and private key. The public key is available to anyone, even to someone who might want to intercept the encoded message. However, as one might guess, the private key is known only by its owner. A message is encrypted using the person's public key, and the only way to decrypt the message is using the intended recipient's private key, which again, is known only to them.

Conclusion

While the invention of the Hill Cypher was a leap forward in encryption in 1929, in today's world it does not provide the type of security required to send anything remotely private or confidential. It is susceptible to numerous different attacks, and once the cypher is cracked, messages can be intercepted and sent along without alerting either party. An effective solution to this problem is the use of RSA encryption, which is extremely secure.