Cryptography

What is Cryptography?

Cryptography is the discipline of encoding and decoding messages. It has been employed in various forms for thousands of years, and, whether or not you know it, is used frequently in our daily lives. Encryption is used to keep our data safe on the Internet, when we use the ATM, and in many other everyday activities. There are two main categories of cryptographic procedures. A **code** works by replacing whole words or phrases with others, at the level of meaning; for example, when a parent substitutes one word for another in front of his or her child. A **cipher** works by transforming and replacing individual letters.

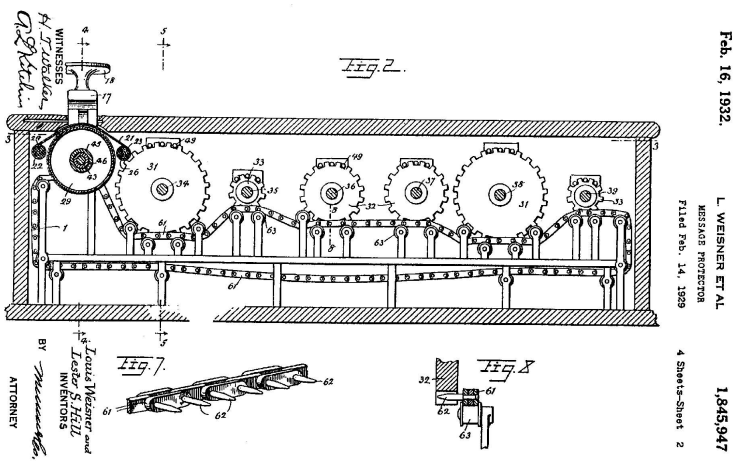
Before we go deeper into the details of cryptography, there is some associated terminology that you should be familiar with. The **alphabet** is the characters that the message will be written in. **Plaintext** or **clear text** is the original message in readable form. We will refer to encoding the message as **encryption** or **enciphering**. The **key** or **password** is the information that is used in the encryption process and is the basis of security for a code. These keys are usually kept private, and the **inverse key** (used to decipher the encoded message) is given to the person needing to decode the message. The encrypted and hard-to-read message is called **ciphertext** or an **encrypted message**. Using the inverse key to turn the ciphertext back into plaintext is referred to as **decrypting** or **deciphering**. If you are able to decrypt the message without being told the inverse key, we call that **cracking** the code.

What is a Hill Cipher?

A Hill cipher is a type of **polygraphic** cipher, where plaintext is divided into groups of letters of a fixed size and then each group is transformed into a different group of letters. A Hill cipher accomplishes this transformation by using matrix multiplication. It was one of the first practical applications of linear algebra to polygraphic ciphers. Hill ciphers were first described by their creator Lester Hill in 1929 in *The American Mathematical Monthly*, and he wrote another article about them in 1931.

Hill ciphers are applications of linear algebra because a Hill cipher is simply a linear transformation represented by a matrix with respect to the standard basis. Groups of letters are represented by vectors. The domain of the linear transformation is all plaintext vectors, while the codomain is made up of all ciphertext vectors. Matrix multiplication is involved in the encoding and decoding process. And when trying to find the inverse key, we will use elementary row operations to row reduce the key matrix in order to find its inverse in the standard manner.

Hill and Louis Weisner also had an idea to build a machine that would mechanically implement a Hill cipher. They named it the Message Protector and patented it. It operated on blocks of six letters and did all of the arithmetic using gears and pulleys. The diagram below is from their patent application. [1]



How to Use a Hill Cipher

A necessary part of Hill ciphers is modular arithmetic. If you would like to learn more about modular arithmetic, click here for a short tutorial.

Encryption

First, we need to specify our alphabet. We will use the standard 26-letter English alphabet, with the characters "?", ".", and "_" (blank space) appended to the end, giving us an alphabet size m=29. We add these characters to make m=29 (a prime number) because working with an alphabet of prime size gives us a guarantee that our key matrix will be invertible modulo m, no matter what matrix we choose. We are able to work with an alphabet size that is non-prime, but it makes the cipher much more difficult to work with, and we have to be careful to pick a key matrix that is invertible.

We will represent each character in our alphabet with a positive integer 0-28 according to the table below. If you wanted to make the cipher more secure, you might scramble the order of the characters before assigning numbers.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	.	?	_
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28

Our message will consist of the plaintext "LINEAR_ALGEBRA_IS_AWESOME." We will follow a simple Hill 2-cipher example throughout, which means that we will divide our plaintext up into vectors of length two (ie two characters). If our message has odd length, we would just append

"x" (or any other alphabet character) to the end of the message. Our Hill 2-cipher will have a key matrix of $A = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}$. We choose the values for our key matrix from the range 0-28. The cipher would be more secure with a longer vector length, but for simplicity's sake, we will just be working with small vectors. (Note: The algorithm is the same for larger keys, there are just more computations involved)

We will now take our plaintext message and, using our table, convert it to its numerical equivalent so that we can encode it. We will also break it up into row vectors of length two.

$$11 \quad 8 \mid 13 \quad 4 \mid 0 \quad 17 \mid 28 \quad 0 \mid 11 \quad 6 \mid 4 \quad 1 \mid 17 \quad 0 \mid 28 \quad 8 \mid 18 \quad 28 \mid 0 \quad 22 \mid 4 \quad 18 \mid 14 \quad 12 \mid 4 \quad 26$$

Now, to encrypt our message, we will take the transpose of each vector and premultiply by A in order to apply the linear transformation.

$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} 11 \\ 8 \end{pmatrix} = \begin{pmatrix} 46 \\ 84 \end{pmatrix}$$
$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} 13 \\ 4 \end{pmatrix} = \begin{pmatrix} 38 \\ 72 \end{pmatrix}$$
$$\vdots$$

This gives us the numerical message:

$$46 \quad 84 \mid 38 \quad 72 \mid 51 \quad 85 \mid 56 \quad 112 \mid 40 \quad 74 \mid 11 \quad 21 \mid 34 \quad 68 \mid 80 \quad 152 \mid 120 \quad 212 \mid 66 \quad 110 \mid 62 \quad 106 \mid 64 \quad 116 \mid 86 \quad 140$$

But wait, these numbers don't appear to correspond to any of the letters in our table! This is where modular arithmetic comes in. We will take all of the values modulo 29 in order to get numbers that correspond to our starting alphabet.

$$17 \quad 26 \mid 9 \quad 14 \mid 22 \quad 27 \mid 27 \quad 25 \mid 11 \quad 16 \mid 11 \quad 21 \mid 5 \quad 10 \mid 22 \quad 7 \mid 4 \quad 9 \mid 8 \quad 23 \mid 4 \quad 19 \mid 6 \quad 0 \mid 28 \quad 1$$

Therefore, our ciphertext message is "R.JOW??ZLQLVFKWHEJIXETGA_B". We now have our encoded message.

Decryption

The next step would be to decode the message. If we are given the key matrix, this is pretty easy. All we need to do is find the inverse of the key because the inverse transformation is the cipher whose matrix with respect to the standard basis is the inverse of A. Modular arithmetic will be useful here too.

To find the inverse of our key matrix, we will now need to do row operations modulo 29. We will need to refer to our table of inverses in order to correctly put our matrix in reduced row echelon form.

$$
\begin{aligned}
\left(A \,\middle|\, I \right) &= \left(\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 4 & 5 & 0 & 1 \end{array} \right) && (1)\\
&\leftrightarrow \left(\begin{array}{cc|cc} 30 & 45 & 15 & 0 \\ 4 & 5 & 0 & 1 \end{array} \right) && (2)\\
&\equiv \left(\begin{array}{cc|cc} 1 & 16 & 15 & 0 \\ 4 & 5 & 0 & 1 \end{array} \right) \pmod{29} && (3)\\
&\leftrightarrow \left(\begin{array}{cc|cc} 1 & 16 & 15 & 0 \\ 0 & -49 & -60 & 1 \end{array} \right) && (4)\\
&\equiv \left(\begin{array}{cc|cc} 1 & 16 & 15 & 0 \\ 0 & 28 & 27 & 1 \end{array} \right) \pmod{29} && (5)\\
&\leftrightarrow \left(\begin{array}{cc|cc} 1 & 16 & 15 & 0 \\ 0 & 784 & 756 & 28 \end{array} \right) && (6)\\
&\equiv \left(\begin{array}{cc|cc} 1 & 16 & 15 & 0 \\ 0 & 1 & 2 & 28 \end{array} \right) \pmod{29} && (7)\\
&\leftrightarrow \left(\begin{array}{cc|cc} 1 & 0 & -17 & -448 \\ 0 & 1 & 2 & 28 \end{array} \right) && (8)\\
&\equiv \left(\begin{array}{cc|cc} 1 & 0 & 12 & 16 \\ 0 & 1 & 2 & 28 \end{array} \right) \pmod{29} && (9)
\end{aligned}
$$

Modular arithmetic can be tricky, so in case you didn't follow what just happened, we are going to walk through it now. Step (1) is our augmented matrix that we will use to find the inverse of A. In Step (2), we multiplied the top row by 15 because, looking at our table, we see that 15 is the multiplicative inverse of 2, our current pivot. Step (4) is the traditional elimination step. In Step (6), we followed the same logic that we used in Step (2) and multiplied the bottow row by 28, the multiplicative inverse of 28. Step (8) is again the standard elimination step. In Steps (3), (5), (7), and (9) all we did was take the matrix modulo 29 to keep the numbers easier to work with (though this step can be performed just once at the end if you would like).

This gives us an inverse key $A^{-1} = \begin{pmatrix} 12 & 16 \\ 2 & 28 \end{pmatrix}$

Now, we take the numerical representation of our ciphertext, divide it up into vectors of length two, and multiply the inverse key by the transpose of each vector.

$$\begin{pmatrix} 12 & 16 \\ 2 & 28 \end{pmatrix}\begin{pmatrix} 17 \\ 26 \end{pmatrix} = \begin{pmatrix} 620 \\ 762 \end{pmatrix}$$
$$\begin{pmatrix} 12 & 16 \\ 2 & 28 \end{pmatrix}\begin{pmatrix} 9 \\ 14 \end{pmatrix} = \begin{pmatrix} 332 \\ 410 \end{pmatrix}$$
$$\vdots$$

This gives us:

$$620 \quad 762 \mid 332 \quad 410 \mid 696 \quad 800 \mid 724 \quad 754 \mid 388 \quad 470 \mid 468 \quad 610 \mid 220 \quad 290 \mid 376 \quad 240 \mid 192 \quad 260 \mid 464 \quad 660 \mid 352 \quad 540 \mid 72 \quad 12 \mid 352 \quad 84$$

Again, we find ourselves in the situation where our numbers don't appear to correspond to any of the letters in our alphabet. Modular arithmetic comes to the rescue again! Taking all of the numbers modulo 29 gives us

$$11 \quad 8 \mid 13 \quad 4 \mid 0 \quad 17 \mid 28 \quad 0 \mid 11 \quad 6 \mid 4 \quad 1 \mid 17 \quad 0 \mid 28 \quad 8 \mid 18 \quad 28 \mid 0 \quad 22 \mid 4 \quad 18 \mid 14 \quad 12 \mid 4 \quad 26$$

If we refer again to our original table, we see that this corresponds to "LINEAR_ALGEBRA_IS_AWESOME.", and our message has come full circle.

Implementing a General Hill n-cipher

Now that we have walked through an example to give you an idea of how a Hill cipher works, we will briefly touch on how you would implement a Hill cipher for a generic n-by-n key matrix with vectors of length n.

1. Separate the plaintext from left to right into some number k of groups of n letters each. If you run out of letters when forming the final group, repeat the last plaintext letter as many times as needed to fill out the final group of n letters.
2. Replace each letter by the corresponding number of its position (from 0 through m-1) in the alphabet to get k groups of n integers each.
3. Reshape each of the k groups of integers into an n-row column vector and in turn premultiply each of those k column vectors by A and take the result modulo m.
4. After arranging all k of the resulting column vectors in order into a single vector of length k·n, replace each of the entries with the corresponding letter of the alphabet.

The result is the ciphertext corresponding to the original plaintext.

When decoding the message, follow the same algorithm but substitute in A^{-1} for A and switch the words ciphertext and plaintext. [1]

For larger n, this algorithm can be fairly easily implemented on the computer.

Breaking a Hill Cipher

Hill ciphers are difficult to break because changing just a few letters in the plaintext can result in a dramatic change in the ciphertext. Another strength of a Hill cipher is that as you increase the size of the groups of letters, you very quickly nullify any usefulness of **frequency analysis** (using probabilistic guessing based on the frequency of occurences of certain letters and phrases in the English language) in breaking the cipher. One major disadvantage, though, is that it is extremely susceptible to what is called a **plaintext attack**. This means that if you manage to obtain enough of the plaintext along with the ciphertext, you can discover the decoding matrix simply by solving a system of linear equations. To help counteract this, you can apply another type of cipher, such as a Vigenère cipher, to the result of the Hill cipher, thus doubly encrypting and scrambling the message.

There is a formal theorem that tells us how to crack a Hill cipher:

Suppose the length m of the alphabet is a prime. Let $p_1,p_2,...,p_n$ be n plaintext vectors for a Hill n-cipher having (unknown) key matrix A, and let $c_1,c_2,...,c_n$ be the corresponding ciphertext vectors. Suppose these plaintext vectors are linearly independent (that is, we have enough plaintext). Form the matrix $P = \left(\vec{p_1} \,\middle|\, \vec{p_2} \,\middle|\, \dots \,\middle|\, \vec{p_n} \right)$ having the plaintext vectors as its columns, and the matrix $C = \left(\vec{c_1} \,\middle|\, \vec{c_2} \,\middle|\, \dots \,\middle|\, \vec{c_n} \right)$ having the ciphertext vectors as its columns. Then the same sequence of elementary row operations that reduces C^T to the identity matrix I reduces P^T to the transpose $(A^{-1})^T$ of the inverse key matrix A^{-1}.

This may sound complicated, but the procedure is actually a lot simpler than it seems (assuming that we are able to determine the size n of the key matrix and the length m of the alphabet). Given enough plaintext and ciphertext (we must capture plaintext and ciphertext messages of at least length n^2), we create the matrices C and P as described above. Then we use row reduction modulo m on the matrix $\left(C^T \,\middle|\, P^T \right)$. If the reduction can be completed successfully, and the resulting matrix is of the form $\left(I \,\middle|\, X \right)$ for some X, then the ciphertext vectors are linearly independent, the key matrix A is invertible, and $X=(A^{-1})^T$. Hence $A^{-1}=X^T$.[1]

We will now go through a simple example to demonstrate how this would really work. Consider the plaintext "MATH" and the corresponding ciphertext "YTBY" where n=2. This gives us plaintext vectors corresponding to $\vec{p_1} = \begin{pmatrix} 12 \\ 0 \end{pmatrix}$, $\vec{p_2} = \begin{pmatrix} 19 \\ 27 \end{pmatrix}$ and ciphertext vectors that are

$\vec{c_1} = \begin{pmatrix} 24 \\ 19 \end{pmatrix}$, $\vec{c_2} = \begin{pmatrix} 1 \\ 24 \end{pmatrix}$. We can therefore construct the matrices $P = \left(\vec{p_1} \mid \vec{p_2} \right) = \begin{pmatrix} 12 & 19 \\ 0 & 7 \end{pmatrix}$ and $C = \left(\vec{c_1} \mid \vec{c_2} \right) = \begin{pmatrix} 24 & 1 \\ 19 & 24 \end{pmatrix}$, in accordance with the theorem. Thus $\left(C^T \mid P^T \right) = \left(\begin{array}{cc|cc} 24 & 19 & 12 & 0 \\ 1 & 24 & 19 & 7 \end{array} \right)$. Like we did earlier, we will now row reduce the matrix modulo 29, giving us $\left(\begin{array}{cc|cc} 1 & 0 & 12 & 2 \\ 0 & 1 & 16 & 28 \end{array} \right)$. From this, we can see that $(A^{-1})^T = \begin{pmatrix} 12 & 2 \\ 16 & 28 \end{pmatrix}$. Therefore $A^{-1} = \begin{pmatrix} 12 & 16 \\ 2 & 28 \end{pmatrix}$. We can then use this inverse key to decrypt subsequent messages.

The most important idea to take away from this is that given enough plaintext and ciphertext, we can break a Hill cipher simply by solving a system of linear equations.

Examples

JavaScript Example of a Hill Cipher [2]

This is a JavaScript implementation of a Hill Cipher. The example here is restricted to a 2x2 case of the Hill cipher. The alphabet for this example is A-Z (ie m=26).

The key should be input as 4 numbers, e.g. "5 17 4 15" (without quotes). These numbers will form the key (top row, bottom row).

Plaintext

linear algebra is awesome

key = 5 17 4 15

v Encrypt v ^ Decrypt ^

Ciphertext

JavaScript Example of Cracking a Hill Cipher [2]

For this example, you get to try your hand at cracking a Hill 2-cipher. You can enter in whatever text you would like in the "Plaintext" box (Note: a shorter message requires fewer calculations to crack). When you click "Start," the computer will generate a random key and give you the ciphertext associated with that key. Your goal is to use this information to crack the cipher and discover the inverse key. If you think you've discovered the inverse key, you can click "Display Inverse Key," and the computer will tell you the inverse key that it generated. We will use the same alphabet and convention for representing the inverse key as in the previous example.

If you would like a refresher on how to break a Hill cipher, click here.

Start

Plaintext

linear algebra is awesome

Ciphertext

Display Inverse Key

Inverse Key

Mathematica Example of a Hill Cipher [3]

Here is an example of the implementation of a Hill cipher in Mathematica.

If you can't see the video or would like to download the Mathematica application to try, you can find it here.

Fun Facts

The term **cryptography** comes from the Greek words *kryptos* (κρνπτοσ), meaning hidden or secret, and *graphia* (γραφια), meaning writing.

The earliest known use of cryptography dates back to approximately 4500 years ago and is found carved into monuments from Egypt's Old Kingdom in non-standard hieroglyphs.

The National Security Agency is reportedly the largest employer of mathematicians in the world. The NSA is the United States government agency that deals with cryptography.

Other Resources

Hill Cipher

Cryptography in an Algebraic Alphabet ~ Lester Hill

Concerning Certain Linear Transformation Apparatus of Cryptography ~ Lester Hill

Hill Cipher Paper

References

[1] An Introduction to Cryptography

[2] Source code from JavaScript Example of a Hill Cipher

[3] Wolfram Mathematica Hill Cipher Example

The Encryption of your Private Network

Hill Ciphers and Modular Linear Algebra

Last modified *Mon 14 December 2009*.