The Number of Possible Keys for the Hill Cipher

Exeter Mathematics School

EXETER MATHS SCHOOL
FREEDOM FOR THINKERS

1 Introduction

This project is an investigation of the Hill cipher, particularly the keys used to encode and decode text with the Hill cipher. The Hill cipher is an algorithm used to encode text with the intention that sensitive information cannot be read by the wrong people. We will find a formula to calculate the number of possible keys for the Hill cipher and then prove this result. We will also discuss how to make the Hill cipher more secure and harder to decode without knowledge of the keys used to encode a particular message.

The Hill cipher [1] uses matrix multiplication to map the plaintext (text to be encoded) onto the ciphertext (text which has been encoded) using the key matrix, $\boldsymbol{A}$ (1). The plaintext and ciphertext are stored in vectors, $\boldsymbol{P}$ and $\boldsymbol{C}$ respectively, which have the same number of rows as the key matrix. Each element of $\boldsymbol{A}$, $\boldsymbol{P}$ and $\boldsymbol{C}$ is in the ring of integers modulo m [2][3], $\mathbb{Z}_m = \{0, 1, 2, \ldots, m-2, m-1\}$. In other words, $\mathbb{Z}_m$ is the set of the smallest positive remainders when dividing the set of integers by m. These integers are used to represent each character in the set of characters being used and therefore m is the order of the character set.

$$\begin{gathered}\boldsymbol{AP} \equiv \boldsymbol{C} \pmod{m} \\ \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,n} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} \equiv \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \pmod{m} \\ \begin{bmatrix} a_{1,1}p_1 + a_{1,2}p_2 + \ldots + a_{1,n}p_n \\ a_{2,1}p_1 + a_{2,2}p_2 + \ldots + a_{2,n}p_n \\ \vdots \\ a_{n,1}p_1 + a_{n,2}p_2 + \ldots + a_{n,n}p_n \end{bmatrix} \equiv\end{gathered} \tag{1}$$

A common example is the letters from A to Z where $\mathrm{A} = 0, \mathrm{B} = 1, \ldots, \mathrm{Y} = 24, \mathrm{Z} = 25$ as in table 1. In this case $m = 26$ as there are 26 letters.

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

Table 1: The Values of each Letter

2 Methodology

To decode text encoded using the Hill cipher the key matrix must be invertible which can be seen in equation 2.

$$\begin{aligned}\boldsymbol{AP} &= \boldsymbol{C} \pmod{m} \\ \boldsymbol{P} &= \boldsymbol{A}^{-1}\boldsymbol{C} \pmod{m}\end{aligned} \tag{2}$$

For the matrix to be invertible in modular arithmetic then the determinant needs to be invertible. In the reals, the inverse of the determinant is simply the reciprocal of the determinant, however, in modular arithmetic the inverse of a, denoted a^{-1}, can be defined by the equation $aa^{-1} \equiv 1 \pmod{m}$. This equation is only satisfied when a and m are coprime or in other terms $\gcd(a, m) = 1$. This is due to Bezout's identity [4] which states that when $ax + by = \gcd(a, b)$ there exists integer solutions for x and y. The expression $aa^{-1} \equiv 1 \pmod{m}$ can be re-written as $aa^{-1} = 1 + km$ where k is an integer. We can rearrange to give that $aa^{-1} - km = 1$ which implies that a is the same for both formulae, $b = m$, $x = a^{-1}$, $y = -k$ and therefore $\gcd(a, m) = 1$ in order for there be integer solutions for x and y and hence for a^{-1} to exist.

A group of invertible matrices is known as a general linear group [5] and therefore the number of invertible matrices is the order of the general linear group. A computer program written in python was used to count how many matrices have determinants which satisfy the conditions for an inverse matrix to exist. A summary of the results generated by the program can be seen in table 2. There are no results for higher values of n because there are m^{n^2} unique matrices with entries from $\mathbb{Z}_m$ as there are n^2 items in the matrix and each item can take m different values. As n increases the number of matrices to check becomes very large and takes too long to compute.

m \ n	1	2	3
2	1	6	168
3	2	48	11232
4	2	96	86016
5	4	480	1488000
6	2	288	1886976
7	6	2016	33784128
8	4	1536	44040192
9	6	3888	221079456
10	4	2880	249984000

Table 2: The Order of $\mathrm{GL}_n(\mathbb{Z}_m)$

It may be noted that the first column is Euler's totient function of m or the number of numbers which are coprime to m. This makes sense because the determinant of a 1×1 matrix is simply the number in the matrix and therefore there are m possible matrices. Because the determinant has to be coprime to m then there are only $\phi(m)$ matrices whose determinants are coprime to m giving the number of invertible matrices for the first column. Each item in the second column is divisible by m and also the item in the first column with the corresponding value of m. Dividing by these numbers gives the sequence 3, 8, 12, 24, 24, 48, 48, 72, 72, ... which is the sequence A007434 [6] in the OEIS (On-Line Encyclopedia of Integer Sequences). Repeating a similar process on the third column by dividing the third column by m^2 and then the column $n = 2$ in the same way gives the sequence 7, 26, 56, 124, 182, 342, 448, 702, 868, ... which is A059376 [7] in the OEIS. These sequences are generated by Jordan's totient function [8], J_k, a generalisation of Euler's totient function as $J_1(x) = \phi(x)$. Jordan's totient function is used to calculate the number of k-tuples which are coprime to the modulus, m. A k-tuple is an ordered list of k items and the definition of a k-tuple and an integer being coprime is if the greatest common divisor of the all of the numbers in the $(k+1)$-tuple formed from adding the integer to the end of the k-tuple is 1. The formula for Jordan's totient function is shown in equation 3.

$$J_k(x) = x^k \prod_{p|x} \left(1 - \frac{1}{p^k}\right) \tag{3}$$

The order of the general linear group of $n \times n$ matrices is a multiple of the product of $J_k(m)$ for increasing values of k from 1 to n. When $n = 2$ we need to multiply the product by m to get to the correct value and when $n = 3$ we need to multiply the product by m^3 to get to the correct value. When $n = 4$ and $m = 2$ then $|\mathrm{GL}_4(\mathbb{Z}_2)| = 20160$ which is 2^6 times the product and when $n = 5$ and $m = 2$ then $|\mathrm{GL}_5(\mathbb{Z}_2)| = 9999360$ which is 2^{10} times the product. The exponents are all triangular numbers and therefore it appears that the equation for the order is $m^{0.5n(n-1)}$ times the product which gives equation 4. The proof of this result may be seen in the report for this project. This result also appears in [9] but no proof is given.

$$|\mathrm{GL}_n(\mathbb{Z}_m)| = m^{\frac{n(n-1)}{2}} \prod_{k=1}^{n} J_k(m) \tag{4}$$

3 Conclusion

In order to make the Hill cipher as hard to crack as possible then the values of m and n should be as large as practically possible and m should be prime. This is because when m is prime the only singular matrices are those with a determinant of 0. However, when m is composite all of the matrices whose determinant has a greatest common divisor with m which is not equal to 1 are also singular and therefore the proportion of matrices which are invertible is lower. To make the cipher more secure this percentage should be as large as possible. This is illustrated for the case when $n = 2$ in figure 1.

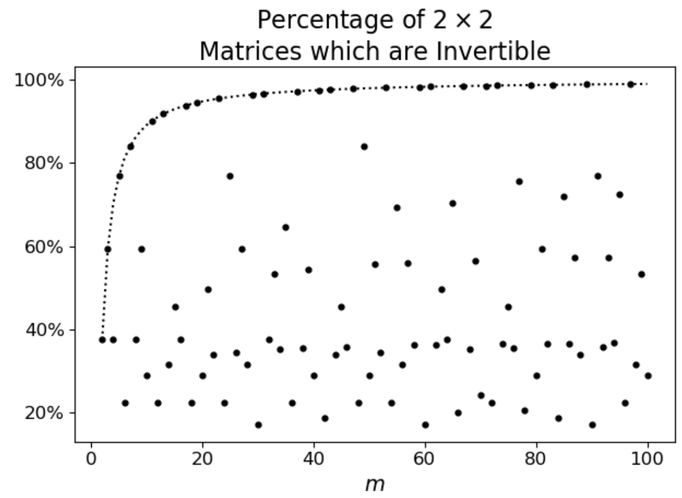

Figure 1: The Percentage of 2×2 Matrices which are Invertible against the Modulus of the Ring of Integers, m

The dotted line represents the number of matrices with a non-zero determinant and all of the points on this line occur when m is prime. Figure 2 shows the same pattern but with the number of matrices as opposed to the percentage (the dotted lines are equivalent).

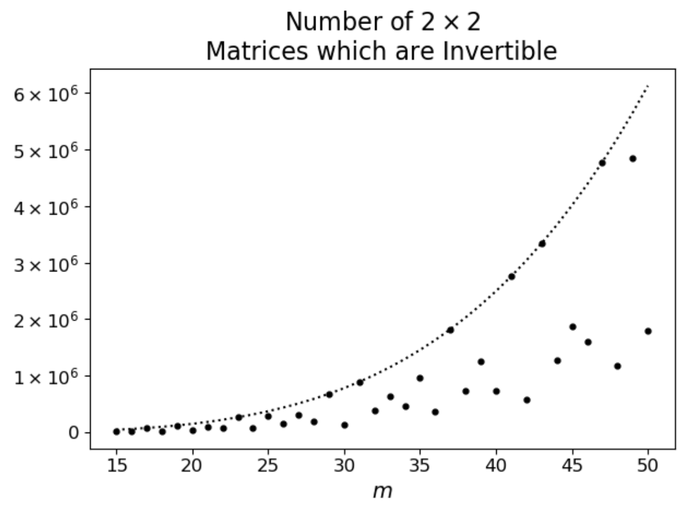

Figure 2: The Number of 2×2 Matrices which are Invertible against the Modulus of the Ring of Integers, m

Comparing figure 2 to figure 3 shows that for 3×3 matrices there are more possible matrices than the same value of m for 2×2 matrices. As well as this the rate of change of the gradient of the dotted line of the 3×3 matrices is greater than the rate of change of the gradient of the dotted line 2×2 matrices. This can be seen on the graph as the line becomes steeper at a greater rate. The number of matrices with a non-zero determinant can be calculated by $m^{0.5n(n-1)}(m-1)(m^2-1)...(m^n-1)$.

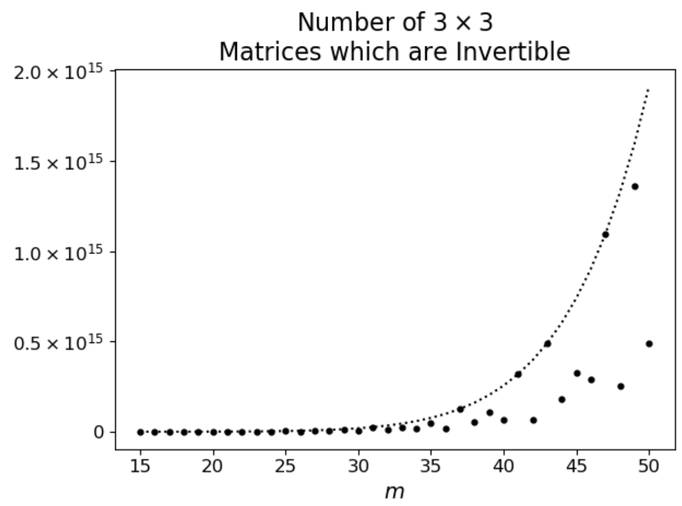

Figure 3: The Number of 3×3 Matrices which are Invertible against the Modulus of the Ring of Integers, m

4 Discussion

In practice m often takes the value of 26 to represent each letter of the alphabet, however, 26 is not prime so the number of possible keys is not as large as possible. By adding extra characters such as a space, full stop and question mark the number of characters can be increased to 29 which means that $m = 29$ and therefore m is prime. This makes the number of possible keys which can be chosen for $n = 2$ much greater; increasing from 157248 when $m = 26$ to 682080 when $m = 29$ which is over 4 times as many keys.

References

[1] Lyons, J. Hill Cipher. [online] Available at: www.practicalcryptography.com/ciphers/classical-era/hill [Accessed 17 May 2018].
[2] Allenby, R. (1991). Rings, Fields and Groups: An Introduction to Abstract Algebra. 2nd ed.
[3] proofwiki.org Definition: Integers Modulo m. [online] Available at: proofwiki.org/wiki/Definition:Integers_Modulo_m [Accessed 11 Sep. 2018].
[4] brilliant.org Bezout's Identity. [online] Available at: brilliant.org/wiki/bezouts-identity [Accessed 21 Dec. 2018].
[5] Terr, D. and Weissein, E. General Linear Group. [online] Available at: mathworld.wolfram.com/GeneralLinearGroup.html [Accessed 17 Feb. 2019].
[6] Sloane, N. A007434. [online] Available at: oeis.org/A007434 [Accessed 22 Dec. 2018].
[7] Sloane, N. A059376. [online] Available at: oeis.org/A059376 [Accessed 22 Dec. 2018].
[8] www.encyclopediaofmath.org (n.d.). Jordan Totient Function. [online] Available at: www.encyclopediaofmath.org/index.php/Jordan_totient_function [Accessed 31 Jul. 2018].
[9] Andrica, D. and Piticari, M. (2003). On Some Extensions of Jordan's Arithmetic Functions. [pdf] Available at: www.emis.de/journals/AUA/acta7/Andrica .pdf [Accesssed 11 Sep. 2018].