Applied Information Technology And Computer Science Vol. 4 No. 1 (2023) 092-108

AITCS

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/aitcs
e-ISSN :2773-5141

A Security Level Comparison of Caesar Cipher, Columnar Transposition Cipher and Row Transposition Cipher in Tamil Messages

Vimaleswari K Veerasingam[1], Nur Ziadah Harun[1]*

[1]Fakulti Sains Komputer dan Teknologi Maklumat,
Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/aitcs.2023.04.01.006
Received 16 September 2023; Accepted 27 May 2023; Available online 30 June 2023

Abstract: Cryptography is a technique that effectively implements privacy and protects confidential information delivered via insecure channels of communication. To present-day, no research has been done to encrypt Tamil letters using the Caesar cipher and Columnar Transposition cipher, and Row Transposition cipher to compare the security level. This research aims to propose a security analysis of the Caesar cipher, and Row Transposition cipher in the Tamil language. This research is to examine the security level of the ciphers and a cryptanalytic attack which is the brute force attack experiment carried out using the Java program. The result will evaluate the security level for the Tamil language through the performance of the ciphers by comparing the complexity of the ciphers to crack the ciphertext using the number of possibilities to crack the cipher.

Keywords: Caesar Cipher, Columnar Transposition Cipher, Row Transposition Cipher, Security Level, Tamil

1. Introduction

Privacy has been the most important prerequisite in communication since humans have been developing new ways of communication and the need to store their messages from other people. As a result, one of the most significant changes in communication today is the importance of cryptography, which has evolved into interdisciplinary technology.

Cryptography was formed from two Greek languages [1]. 'Krypto' means "secret" or "hidden," and 'Graphene" means "writing," thus cryptography refers to "secret writing." Cryptography is known as a technique that succeeds in protecting confidential information transmitted through insecure communication [2]. Cryptography is a set of techniques that will be used to convert readable information into an unreadable format for unauthorized individuals. There are two processes involved in cryptography which are encryption and decryption [3]. The process that encrypts the plaintext to ciphertext is called encryption whereas the process that decrypts the ciphertext to the plaintext is called decryption.

Corresponding author: nurziadah@uthm.edu.my

There is plenty of historical ciphers such as Caesar cipher, Transposition cipher, Playfair cipher, Rail Fence cipher and many more. In this research, we are only focusing on three ciphers which are the Caesar cipher, Columnar Transposition cipher and Row Transposition cipher. Caesar cipher is one of the easiest and oldest ciphers. [4]. Caesar cipher is an example of a substitution method in cryptography. According to [5], a transposition cipher is an example of a permutation cipher. The transposition cipher will permute or change the position of the letters in plaintext.

Nowadays, hackers are always trying to break the cryptographic methods or retrieve keys using modern cryptography and tend to forget about classical cryptography. Hence, as an ancient quote says it's true that "Old is Gold" those techniques help in developing new ideas to sort out problems, thus leading to the discovery of a new perspective of the techniques [6]. Besides, most historical cipher applications are only applicable to English alphabets to encrypt and decrypt. To present, there is only English and Jawi language has been used for research about ciphers. The Tamil alphabets might have a different outcome compared with English and Jawi characters as different characters have different characteristics.

Therefore, the project aims to study the encryption of Tamil Language messages using Caesar cipher, Columnar Transposition cipher and Row Transposition cipher. The objective of the research is to conduct security analysis on Caesar cipher, Columnar Transposition cipher and Row Transposition cipher in Tamil Language ciphertext and evaluate the security level by comparing the complexity of cracking the Tamil Language ciphertext.

This study focuses on doing security analyses on three historical ciphers: the Caesar cipher, the Columnar Transposition cipher, and the Row Transposition cipher. For this study, Tamil language messages were used. To encode and decode Tamil messages, this research utilizes JAVA coding for Caesar cipher, Columnar Transposition cipher, and Row Transposition cipher. The encryption technique takes the message in Tamil as input. For this study, the size of the key and the key itself will be produced at random.

The rest of the paper is organized as follows: Section 2 discussed related work cryptography, cryptanalysis, and other research contribution. Then, Section 3 described the detection methodology which has been used in this paper. Section 4 explained the result of the experiment and finally, Section 5 discussed the conclusion and future work.

2. Related Work

2.1 Historical Ciphers

Classical cipher algorithms were designed and used long before public-key cryptography was proposed. It does not involve mathematical concepts in it [7]. The substitution method and transposition method are the two significant methods used in classical ciphers [8]. The substitution method is the replacement of the letters in plaintext with another letter corresponding to it. Caesar cipher is the cipher that uses the substitution method in historical ciphers [2]. The transposition method will permute or change the position of the letters in the plaintext. A transposition cipher is a cipher that uses the transpose method in historical ciphers [2]. In this research, the historical cipher is more focused on the Caesar cipher, Columnar Transposition Cipher and Row Transposition cipher.

2.1.1 Caesar Cipher

Caesar cipher is also referred to as Caesar's shift or Caesar's code. King Julius uses this method by shifting 3 characters in the plaintext and the ciphertext [9]. Using the shift key, each letter in the plaintext is replaced by a letter that corresponds to a specific letter. The original encryption key in Caesar cipher is three-character shifts to the right. The Caesar cipher is improved by implementing modulo twenty-six arithmetic (mod 26). Modulo twenty-six represents the total number of characters

in the English alphabet. Modular 26 is used in the encryption and decryption formulas inequation 1 and 2, which refers to the total number of letters in the English alphabet, omitting spaces between characters.

The Caesar cipher encryption method is explained in Equation 1 [10].

$$C = E_n(x) = (x + n)\ mod\ 26 \qquad Eq.\ 1$$

The Caesar cipher decryption method is explained in [10].

$$P = D_n(x) = (x - n) mod\ 26 \qquad Eq.2$$

Where,

C = ciphertext D = decryption
P = plaintext n = number of shifted keys
E = encryption x = alphabets position in plaintext message

2.1.2 Columnar Transposition Cipher

In a columnar transposition cipher, the plaintext is written horizontally on a fixed-width sheet of paper, and the ciphertext is read vertically. Whereas the ciphertext is written on the same width of the paper and the plaintext is read horizontally for the decryption process of the columnar transposition cipher [3]. The plaintext has been padded to fit inside a rectangular box. This is identified as a regular columnar transposition. If there are characters that are left blank in the transposition cipher it's called irregular transposition [11].

2.1.3 Row Transposition Cipher

The plaintext for a row transposition cipher will be written horizontally, and the ciphertext will be written horizontally as well, but it follows the arrangement from the given key given for the encryption.

The purpose of choosing the Caesar cipher, Columnar Transposition cipher and Row Transposition cipher for this research is because all the ciphers using a key and the similar key will be used to decrypt the message. Therefore, it makes us easy to use the technique. It also involves a safe method to transfer the key from one to another.

2.2 Tamil Language

The Tamil language is one of the international languages. The Tamil language is widely spoken among people worldwide. The Tamil language is a part of the Dravidian language family, specifically the South Dravidian languages, which include Kannada, Malayalam, and Tulu [12]. The Tamil alphabet is also known as an abugida script where consonant-vowel sequences are represented as a unit. In the Tamil language, there are a total of 247 alphabets which consists of 12 vowels (Uyir-Eluthukkal (□□□□□ □□□□□□□□□□□□□)), 18 consonants (Mei-Eluthukkal (□□□□ □□□□□□□□□□□□□)) and a special character called āytam (□) [13]. Āytam (□) is also known as āyutha eluttu (□□□□ □□□□□□□□) which is classified as being neither a consonant nor a vowel. The combinations of the vowels and consonants are the remaining alphabets, and they are called consonantal vowels which is Uyir Mey Ezhuthukal (□□□□□□□□□ □□□□□□□□□□□□□). Table 1 shows the list of Tamil characters.

The Tamil language is very sensitive. This is because in most Tamil alphabets it will have the same pronunciation but with tone and it will change into a new word. For example, □□ and □□ are pronounced as “RI” and “RRI” respectively. If we change these consonantal vowels in and word, it will give a new meaning to it. Example: -

- □□□ “KARI” means ash

- □□□ “KARRI” means curry (a type of food)

Table 1: Tamil Characters

	□	□	□	□	□	□	□	□	□	□	□	□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□
□□	□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□	□□

2.3 Cryptanalysis

Cryptanalysis is the process of decrypting the ciphertext. Cryptanalysis will analyse the codes, ciphers or encrypted text [14]. The goal of cryptanalysis is to crack the cryptography of information security systems. The following are examples of cryptanalytic attacks [15]:

i. Cipher text-only attack: The attacker has a string of cypher text.
ii. Known plain text: The attacker has a string of plain text, P, and the corresponding cypher text, C.
iii. Selected plain text: The attacker has gained temporary access to the encryption machinery. As a result, he or she can select a plain text string, P, and construct the corresponding cypher text string, C.
iv. Chosen cypher text: The attacker has gained temporary access to the decryption machinery. As a result, he or she can select a cypher text string, C, and construct the corresponding plain text string, P.
v. Brute Force Attack: A brute force attack is a method of breaking a cypher that involves attempting every possible key. Because of the extensive search, the brute force attack is the most expensive.

In this research, the experiment is focused on brute-force attack. A brute-force attack is an attempt to crack a password by systematically trying every possible combination of letters, numbers, and symbols until you find the one that works. Brute force is an attack that employs the "trial and error" method of guessing [16]. In this research, the cryptanalysis focuses on three ciphers.

i. Caesar Cipher: The trial-and-error method or brute-force is used in the cryptanalytics process for the Caesar cipher. The method procedure may begin by picking a specific section of the ciphertext. The shift key must be discovered throughout the decoding procedure. The shift key has a number range of one to twenty-six which is the range of the alphabet in the English language. To decipher the ciphertext, the decryption algorithm tries each shift key in the message one by one. Shift back the ciphertext to get the original message when the shift key is identified. The original letter replaces the letter of ciphertext. Each possible decrypted message will result in the readable text during the decryption process. The plaintext is the readable text that we have found. [11].
ii. Columnar and Row Transposition cipher: The block size in the transposition cipher will be guessed to conduct cryptanalytics attack. The ciphertext is divided into blocks of varying lengths based on a block size guess. Each block is treated as a matrix row. The purpose of cryptanalytics is to rearrange the order of each character in the block. The cryptanalytics procedure may select a column with many common characters so that the possibility of breaking the ciphertext is increased. The brute force attack process will continue until the plaintext is found.

There is much research that has been done on cryptanalytics and ciphers. Cryptanalytics attack allow us to know the security elements of each cipher. The security elements in each cipher will help to protect the plaintext in terms of confidentiality. Work by [17] used the Vigenère cipher to cryptanalysis ciphertext to provide an implementation of the cryptanalysis method. The proposed research shows to implement the message without errors using simple algorithms. The study will assist us to develop a better understanding of the mathematical formulas used to analyses secure information systems to find weaknesses and hidden components. The research first processed the Vigenère cipher using three main steps which are obtaining the key, determining the length of the key, and determining the key characters. Firstly, the IC for the keywords will be calculated using the length of 1. To allow the calculation of the ciphertext to be easy, space and symbols will be removed from the ciphertext. The same method will be done for the keyword with a length of 2 to 10. After calculating the average keyword length of IC, the largest IC average will be selected to get the key length. Then, the ciphertext will be divided into groups using the length of the key found and by using English language frequency the frequency of letters will be presented to compare. Next, a similarity chart will be built by shifting right and left manually. The number of shifts needed will be key for the Caesar cipher. The Vigenère cipher will take the character on the top as the keyword. Finally, the Vigenère cipher method will be applied after the key length and key characters are determined. In conclusion, the research will explain the cryptanalysis process in every detail on how this technique can recover the plaintext if the encryption keys are known. It also analyses the methods mathematically on how to decrypt the ciphertext with no errors and to get accurate results.

Work by [11] compare Caesar cipher and Transposition cipher security levels using Jawi messages. The research was divided into three parts where first the Jawi message will be encrypted, and then the cryptanalysis and security analysis will be done. The research first will encrypt a Jawi message using Caesar cipher or Transposition cipher respectively. Then, the ciphertext will be used for the cryptanalytics attack process. At this stage, the Jawi ciphertext will be separated the letter by letter to count the distribution of the Jawi letters. Based on the distribution of the Jawi letters, a histogram will create to compare the frequency of messages. Then, the security level of the ciphers will be examined based on the difficulty in the cryptanalysis of the ciphers respectively. For the cryptanalytics attack of the Caesar cipher, the frequency of the Jawi letters ciphertext will be compared with the frequency of the Jawi letters plaintext. By comparing both frequencies we can conclude which letter represents which letter by this shift key will be found and the message will be revealed. In Transposition cipher, the frequency of the letters will be the same as it changes the position of the letters. The decryption process will be hard if the keyword's length and key letters are unknown. To decrypt, we must represent the keyword's length as kl. Then to identify the number of the row and length of ciphertext we must represent it with r and cl respectively. Therefore, the formula will be $r = \frac{cl}{kl}$. Then, we can write the

ciphertext in column and read the message using the row. Frequency analysis has been conducted as the security analysis for both ciphers. Both ciphers can be cracked using frequency analysis but to determine the secure cipher the ciphers were compared with the number of steps to crack the message. They have made a conclusion where transposition cipher needed more steps to crack the message compared to the Caesar cipher. Hence, the Transposition cipher is safer to use compared to the Caesar cipher.

Work by [18] focuses on cryptanalysis of classic ciphers that will provide a survey of the implementation of cryptanalysis of the method. Three ciphers have been used for this research which are the Caesar cipher, Transposition cipher and hill cipher. The research helps to show to recover a message using a simple algorithm and it will also improve the understanding of the mathematical formulas used to analyse secure information systems to find weaknesses and hidden components. For Caesar cipher and Transposition cipher, a frequency analysis was conducted to crack the ciphertext. In Caesar cipher to find the key, the research has used a formula where the fixed position of highest frequency in ciphertext must with the position of the letter with the highest percentage of repetition of letter frequency whereas for transposition cipher the columns of the ciphertext were rearranged and needed to check whether the arrangement of the row can produce an understandable message. The process will continue until the full ciphertext is changed to a readable message and until the key is obtained. For Hill cipher to perform a ciphertext attack, the crib dragging method is used. The process was done by using the cribs word to get the key. The decryption of the ciphertext was unsuccessful as the key matrix was not invertible. The size of the key matrix was 2*2. To find the correct key matrix, a four linear equation was solved. Then, ciphertext was successful to decrypt using all possible key matrices 2*2. In conclusion, this research explained the techniques to obtain encryption keys and crack the ciphertext using Caesar cipher, Transposition cipher and Hill cipher.

The proposed research for this paper is like the previous research paper in that both are concerned with cryptanalysis and historical cipher. Comparable to [11], the proposed research focuses on Caesar Cipher, Columnar Transposition Cipher, and Row Transposition Cipher. The main distinction between the presented study and previous studies is that we used Tamil as the dataset language, whereas [17, 11, 18] used Jawi and English. For the experiment in this study, a Java program will be run. At the end of the research, the result will be evaluating the secure cipher for the Tamil language through the performance of the ciphers by comparing the complexity of the ciphers to crack the ciphertext using the number of iterations for the decrypted text.

3. Methodology

The framework based on the cipher and the language utilized is discussed in this section. The encryption of Tamil messages, cryptanalysis, and security analysis of the ciphertext are the three key activities that will be included in this research framework. Figure 1 demonstrates a research framework that was modified from [11].

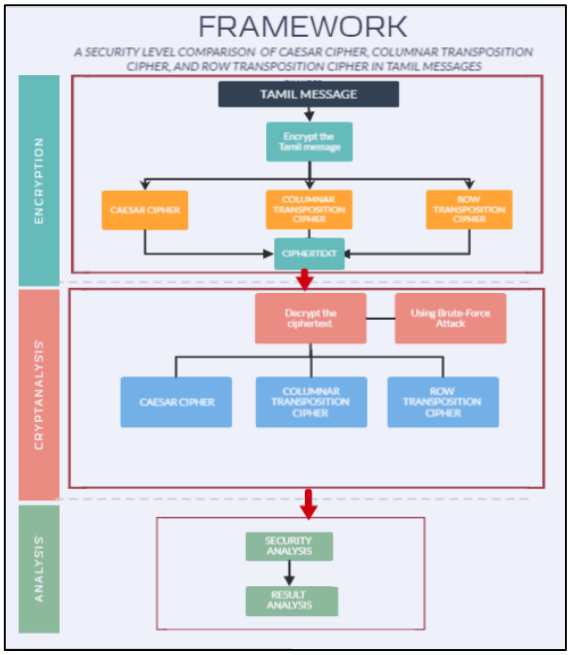

Figure 1: Research framework [11]

3.1 Dataset

Datasets are the Tamil messages used as the input for the study. For this study, a dataset has been created where it consists of a few Tamil messages from Thirukural. Thirukkural is an ancient book of wisdom, the finest classic of the Tamil language, and the distillation of Tamil Nadu's basic brilliance [17]. It was written by Thiruvalluvar and has 1330 couplets or Kurals. The datasets consist of 7 words consist with different lengths. Each words have different meaning and it's a Kural from Thirukural.

3.2 Encryption

The objective of conducting the encryption process is to convert the plaintext by using the secret key to produce ciphertext [8]. The ciphertext message contains all the information contained in the plaintext message, but it is not in a format readable by a person or computer without the right method to decrypt.

3.2.1 Encryption of Caesar Cipher

To begin, the shift key is selected from a list of key values. A number from 1 to 71 is chosen. This is because although Tamil has 247 alphabets, but the main Tamil letters are only 72. The remaining alphabets are produced by adding two of the Tamil alphabets. The initial alphabets in Tamil will be counted as 0 and the last alphabet will be 72, resulting in a range of 1 to 71. [19]. To produce the ciphertext, the key values will be utilized to shift the Tamil letters. The original letter is shifted to the right according to the alphabet's sequence in the substitution procedure. For example, the number 3 is chosen as the key because it was traditionally used for Caesar cipher encryption [9]. The encryption technique for the first nine letters of the Tamil alphabet is shown in Table 3. It also shows that if the key value is 3, then the Tamil alphabet "□" will be replaced by "□" as the ciphertext.

Table 3: The encryption of first 9 letters of Tamil alphabets

Plaintext	□	□	□	□	□	□	□	□	□
Ciphertext	□	□	□	□	□	□	□	□	□

Besides, the encryption can be done by using the formula but we must change the mod number to 72 as there are 72 alphabets in the Tamil language. Therefore, we will still get the same results as the manual shifting way.

$$C = E_n(x) = (x + n)\ mod\ 72$$

$$E_n(x) = (1 + 3) \bmod 72$$

$$= 4 \bmod 72$$

$$= 4$$

$$= \square$$

3.2.2 Encryption of Columnar Transposition Cipher and Row Transposition Cipher

The Tamil message's total amount of characters is determined. Punctuation marks, symbols, and space are all not included in the total amount of characters. In both the Columnar Transposition Cipher and the Row Transposition Cipher, the first part of encrypting Tamil messages is the same. The key, for beginnings, is a number generated at random depending on the key size. The key size determines the block size of the Tamil message. The number 0 is absent from the key. The number of candidate keys created is determined by the key size. Only one of the candidate keys is in use. Divide the total number of characters by the key size to get the remainder. For adding random Tamil characters, there are two options. If the first scenario's remainder is zero, the random Tamil character is not required to add the Tamil message. If the remainder in the second scenario is not zero, the random Tamil character is inserted one by one in the Tamil message until the remainder is zero[11]. The arrangement of the sentences " □□□□ □□□□□□□ " in the column of 4 where there are no remaining boxes is shown in Table 4 and Table 5 shows how " □□□□ □□□□□□□ " is arranged when a random Tamil character '□' is needed to fill the empty space when there is a remainder.

Table 4: If the remainder is zero when the key size is 4

1	2	3	4	KEY
□	□□	□	□	MESSAGE
□	□□	□	□□	

Table 5: If the remainder is not zero when the key size is 4

1	2	3	4	KEY
□□	□□	□	□	MESSAGE
□□	□	□□	□	

The plaintext will be permuted based on the key values after all the empty spaces have been filled. The ciphertext will be determined using the encryption technique. The ciphertext in a columnar transposition cipher is read vertically from the table. The ciphertext will be read horizontally in the Row Transposition cipher.

3.3 Cryptanalysis

In cryptanalysis, the objective is to determine the key size and decryption key to retrieve the plaintext. The encryption method uses the same key size. The decryption key may or may not be the same as the encryption key. Once the key size and key can be recognized and reverse the ciphertext to expose the original message. The person attempting to decrypt the ciphertext must examine the text and attempt to comprehend the contents of the decrypted message.

In this research, the cryptanalysis focusses on three ciphers.

i. Caesar cipher: The Caesar cipher simply substitutes the letter of the original message with another letter. The letters of the Tamil messages substitute to the left, translating the frequency patterns to the left as predicted by the messages. To crack the Caesar cipher, in the cryptanalysis process, the

trial-and-error method is used. The cryptanalysis method begins by selecting a section of the ciphertext. To decipher the ciphertext, the decryption algorithm tries each shift key in the message one by one. Once the shift key can be recognized and reverse the ciphertext to expose the original message. Now, the plaintext will be readable text. During the decryption process, every potential of the decrypted messages can be readable text.

ii. Columnar Transposition cipher and Row Transposition cipher: The positions of Tamil characters and the addition of random Tamil characters are altered. The trial-and-error method is used in the cryptanalysis of the Columnar Transposition cipher and Row Transposition Cipher encryption. The plaintext position is permuted using the Columnar Transposition cipher and Row Transposition Cipher encryption. To rearrange the place of ciphertext, the trial-and-error method is utilized. There are two stages to the decryption process. The Tamil ciphertext is divided into blocks in the first stage. The block size tries to start at one and rise in increments of one. The next step is to try every available key. Both stages must test the key size and key. In the decryption procedure, the trial-and-error method guesses the key size and key. Reverse the ciphertext to reveal the original message once the key size and key have been identified. In the cryptanalysis process, for columnar transposition cipher and row transposition cipher, the ciphertext that has been changed to the plaintext will be read horizontally from the table the trial-and-error method is used. Now, the plaintext will be readable text.

3.4 Performance Evaluation Metric

Selecting an appropriate evaluation metric is critical for discriminating and obtaining the best comparison. In this study, we make a comparison based on the total number of brute force attacks for cracking the ciphertext.

i. Number of Brute Force Attacks

The number of brute force attacks t required to find an optimal solution for a given accuracy determines the overall computational effort and algorithm performance. In this study, the number of brute force attacks is used to get all the possibilities of cracking the ciphertext. The Equation 3 is used for the brute force attacks is the:

$$x_{n+1} = f(x_n),\ n \geq 1 \qquad Eq.\ 3$$

Where, (x_n) is the *n*th approximation or brute force attacks while, $n + 1$ = next brute force attacks of x

3.5 Security Analysis

The cryptanalysis process was used to find out the possible plaintext from the ciphertext that was produced from the encryption process by Caesar cipher, Columnar Transposition cipher and Row Transposition Cipher. The analysis of the security in the ciphers is evaluated by the complexity of decrypting the message from ciphertext to original text. The analysis will look at which encryption is more secure than the others by seeing which one requires the fewest attempts to decrypt the message. If there are fewer attempts, the encryption is less secure than the other.

3.6 Result Analysis

The security analysis report is provided after this study. The report details the Caesar cipher, Columnar Transposition Cipher, and Row Transposition Cipher security levels. The security level decides which cipher makes the Tamil plaintext message simple to recover, suggesting that the cipher is less safe. The results will also be used to decide which encryption is best for Tamil messages.

3.7 Experimental Design

The experiment starts when the Tamil language messages are used for the encryption process. The encryption process will use Caesar cipher, Columnar Transposition cipher and Row Transposition cipher in a tool that compromises the Tamil Language. At the end of the first process, a ciphertext will be produced. Then it will pass through the Brute-force attack process with the ciphertext that has been produced from the encryption process. The brute-force attack of the ciphertext will lead to revealing the original Tamil message which is readable. In brute force attack, the tool that compromises the Tamil Language will show the attempts that have been taken to crack the message. Finally, an evaluation will be done to examine the security level of the Caesar cipher, Columnar Transposition cipher and Row Transposition cipher in the Tamil language by using the maximum number of tries to get the correct key for the ciphertext.

In the Tamil language, there are a total of 247 alphabets. In this scenario, since we are using a Java program to do the encryption of the plaintext, there will be only 72 alphabets needed to produce all the 247 alphabets as the Unicode of the Tamil language is only 72 codes for Java programming. All Tamil characters are now encoded according to the Universal Principle of Unicode. In the Unicode character set, Tamil characters range from U+0B80 to U+0BFF [20]. Figure 4 shows the snippet for the Tamil language Unicode used for the program. Figure 5 shows the Unicode Version 10.0 for Tamil Characters.

```
14    static char[] alphabets = {'\u0b82', '\u0b83', '\u0b85', '\u0b86', '\u0b87', '\u0b88', '\u0b89', '\u0b8a',
15            '\u0b8e', '\u0b8f', '\u0b90', '\u0b92', '\u0b93', '\u0b94', '\u0b95', '\u0b99', '\u0b9a', '\u0b9c',
16            '\u0b9e', '\u0b9f', '\u0ba3', '\u0ba4', '\u0ba8', '\u0ba9', '\u0baa', '\u0bae', '\u0baf', '\u0bb0',
17            '\u0bb1', '\u0bb2', '\u0bb3', '\u0bb4', '\u0bb5', '\u0bb6', '\u0bb7', '\u0bb8', '\u0bb9', '\u0bbe',
18            '\u0bbf', '\u0bc0', '\u0bc1', '\u0bc2', '\u0bc6', '\u0bc7', '\u0bc8', '\u0bca', '\u0bcb', '\u0bcc',
19            '\u0bcd', '\u0bd0', '\u0bd7', '\u0be6', '\u0be7', '\u0be8', '\u0be9', '\u0bea', '\u0beb', '\u0bec',
20            '\u0bed', '\u0bee', '\u0bef', '\u0bf0', '\u0bf1', '\u0bf2', '\u0bf3', '\u0bf4', '\u0bf5', '\u0bf6',
21            '\u0bf7', '\u0bf8', '\u0bf9', '\u0bfa' };
22
```

Figure 4: Code for the Tamil Language

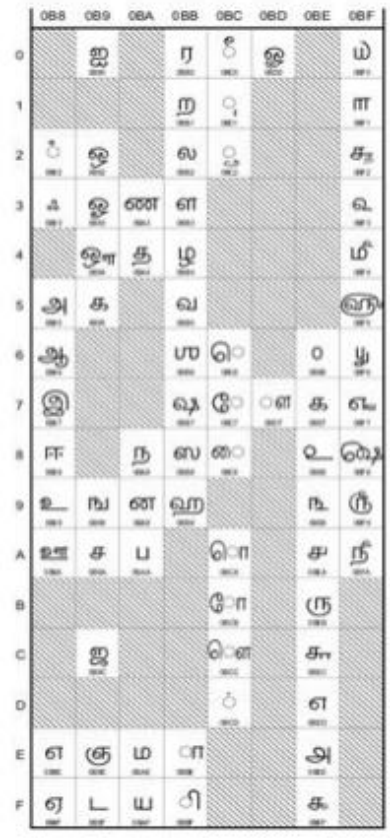

Figure 5: Unicode Version 10.0 for Tamil Characters character [19]

Table 6 depicts the steps involved in the compression of a Tamil word. The word □□□□□□□ seems to have 4 characters, but it is a combination of 7 Unicode characters listed in the Table 6 The Unicode characters combination of the word □□□□□□□ is given below □□□□◌□□□.

Table 6: Combination of Unicode characters for a single Tamil character [20].

Tamil Character in text file	Combination of Unicode characters	Unicode 16 bit
□	□	0B8E
□□	□	0BB4
	□	0BC1
□□	□	0BA4
	◌□	0BCD
□□	□	0BA4
	□	0BC6

To make sure the comparison can be done fairly after the experiment some variables such as program code were kept constant. The manipulated variable in this experiment was the cipher itself. Table 7 shows all the parameter usen in the research.

Table 7: Parameters used for the research

Parameter	Value
Technique	Caesar cipher, columnar transposition cipher and row transposition cipher
Message length/dataset	Range from 3 to 26 characters
Key size for columnar transposition cipher and row transposition cipher	4
Key for encryption and decryption	Randomly generated

4. Results and Discussion

In this experiment phase, the classifier performance evaluation is conducted based on the proposed algorithm, which is Caesar cipher, Columnar Transposition Cipher, and Row Transposition Cipher. In this experiment, the total number of brute forces of possibilities of cracking the ciphertext are tabulated. The same dataset that was used to test for all the ciphers.

4.1 Results and Analysis

Table 8 shows the results for the number of brute forces of the experiment conducted on Caesar cipher, Columnar Transposition cipher and Row Transposition cipher. The results show that the number of brute forces for Caesar cipher, columnar transposition cipher and row transposition cipher for different number of alphabets in the dataset are same respectively. This is due to the constant of the experiment. For the Caesar cipher experiment, the number of alphabets is kept constant which is 72 alphabets. During the cryptanalysis process, the first alphabet is considered as 0, therefore there will be 71 brute force attacks. The number of brute force attacks for columnar transposition cipher and row transposition cipher is same because the key size of the experiment is kept constant which size ‘4’. Figure 5, Figure 6 and Figure 7 show the example results of Caesar Cipher, Columnar Transposition Cipher and Row Transposition Cipher from the program respectively.

Table 8: Results for the number of brute force attacks of the experiment conducted on Caesar cipher, Columnar Transposition cipher and Row Transposition cipher

Number of alphabets in the dataset	The number of brute force attacks		
	Caesar Cipher	Columnar Transposition Cipher	Row Transposition Cipher
3,6,10,13,15,19,26	71	768	768

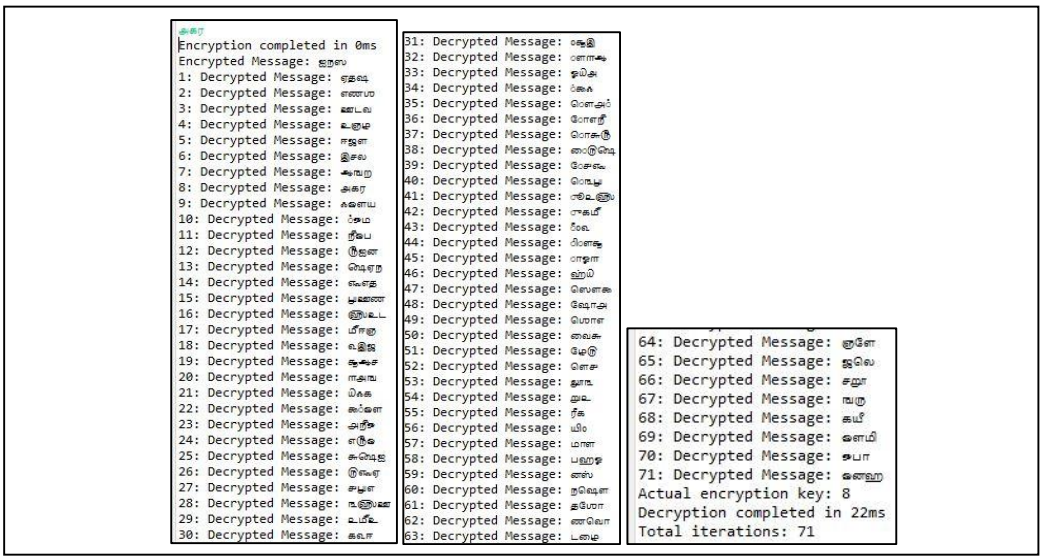

Figure 5: Caser Cipher output from the program

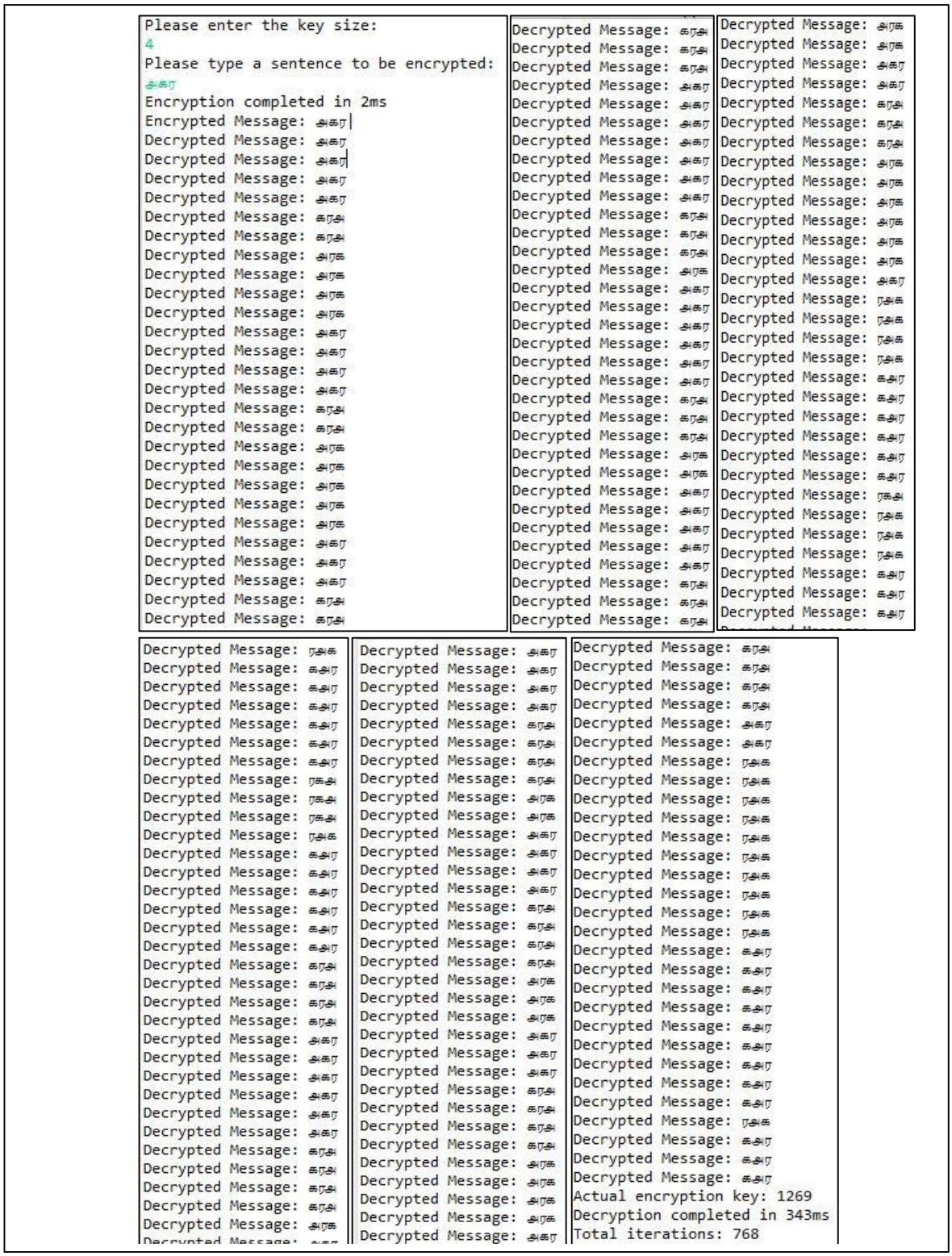

Figure 6: Columnar Transposition Cipher Output from the program

Figure 7: Row Transposition cipher output from the program

Table 9 shows the results for the number of brute force attacks of the experiment conducted on Columnar Transposition cipher and Row Transposition cipher. The results show that the number of brute force attacks for columnar transposition cipher and row transposition cipher for different number of key size but the alphabets in the dataset are same respectively.

Table 9: Results for the number of brute force attacks of the experiment conducted on Columnar Transposition cipher and Row Transposition cipher with different key size

Key Size	Number of brute force attacks	
	Columnar Transposition Cipher	Row Transposition Cipher
3	54	54
4	768	768
5	12500	12500
6	233280	233280

From the experiment, we can find that the higher the key size for the experiment, the more the number of brute force attacks has taken place to the cryptanalysis process. Columnar Transposition cipher and Rows Transposition cipher have roughly the same security level. The only distinction between these two ciphers is the character arrangement in row and column. The historical cipher has a low level of security. The ciphertext generated is decipherable.

From the experiment done, this research evaluates the strength of ciphers on Tamil messages. The results show that the historical cipher has a low level of security. The ciphertext generated is decipherable. The decryption process takes only a few seconds. Columnar Transposition Cipher has a higher security level than Rows Transposition Cipher and Caesar Cipher. This is because the Columnar Transposition ciphers took longer to complete the cryptanalysis process than the Rows Transposition cipher and Caesar cipher. The Caesar cipher consists solely of substituting characters in the same row as plaintext. Rows Columns Transposition cipher and Columns Rows Transposition cipher have roughly the same security level. The only distinction between these two ciphers is the character arrangement in row and column.

A Java program was developed to conduct the encryption and decryption of Tamil text. The program consists of Caesar cipher, Columnar Transposition cipher and Row Transposition cipher. The program was developed to ease the process of encryption and decryption of Tamil text. These programs assist in the analysis to faster the cryptanalysis process. These encrypt and decrypt programs replace the manual method. The step conducted in encrypt program and decrypt program are the same as the step used in the manual method. For Caesar's cipher, the encrypt program takes the plain text and randomly generates the key to generate the ciphertext. The decrypt program takes the ciphertext and takes all the alphabets to substitute to produce all the possibilities of decrypted text and will also produce the correct key. For both transposition ciphers, the encrypt program takes the plaintext, and randomly generates the key to generating ciphertext. The decrypt program takes the ciphertext, and key size to generate all possible keys and corresponding decrypted text. The process to find the correct decrypted text is the same as the manual method. The decrypt program only can provide all the possible decrypted text. The decrypt program cannot find out the correct decrypted text in the program itself. The users need to select the correct decrypted text based on their understanding.

5. Conclusion

As a conclusion, the main objectives of this research have been achieved which is to evaluate the security level of Caesar cipher, Columnar Transposition cipher and Row Transposition cipher in Tamil Language ciphertext by comparing the complexity of cracking the Tamil Language ciphertext the total number of brute force attacks of possibilities of cracking the ciphertext and the total time taken for the process of the classifiers in analyzing the dataset to complete. From the comparison, Caesar cipher is less secure in securing confidentiality in the communication of the Tamil Language. Besides, from this research, we found that Columnar Transposition cipher and Row transposition cipher take a longer time to decrypt the ciphertext. From this, we can say that the transposition cipher is more secure for the Tamil language.

There are some limitations found in this study. The datasets used had a small number of words. The results could have been more accurate if the datasets are more than 100 alphabet. Furthermore, due to time constraints, the cryptanalysis was performed in three ciphers during the study. This study may be able to perform cryptanalysis on other ciphers.

This research can be improvised in future by improving certain aspects of the research. Therefore, some suggestions are:

i. It is suggested that this study be conducted in other languages in the future, such as French, Japanese, and others.

ii. The study should compare the security level of ciphers in other languages using both historical and modern ciphers.

Acknowledgment

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia for its support.

References

[1] J. Hernandez-Castro and G. Avoine, "Cryptanalysis of ubiquitous computing systems," *2016 18th Mediterr. Electrotech. Conf.*, pp. 1–4, 2016.

[2] A. M. Qadir and N. Varol, "A Review Paper on Cryptography," *2019 7th Int. Symp. Digit. Forensics Secur.*, pp. 1–6, 2019.

[3] B. Schneier, "Applied cryptography, second edition : protocols, algorithms,and source code in C," 2015.

[4] S. N. Gowda, "Innovative enhancement of the Caesar cipher algorithm for cryptography," *2016 2nd Int. Conf. Adv. Comput. Commun. Autom.*, pp. 1–4, 2016.

[5] G. S. Wulandari, W. Rismawan, and S. Saadah, "Differential evolution for the cryptanalysis of transposition cipher," *2015 3rd Int. Conf. Inf. Commun. Technol.*, pp. 45–48, 2015.

[6] A. H. Disina, "ROBUST CAESAR CIPHER AGAINST FREQUENCY CRYPTANALYSIS USING BI-DIRECTIONAL SHIFTING.," 2014.

[7] R. Mahendran and K. Mani, "Generation of key matrix for Hill cipher encryption using classical cipher," *2017 World Congress on Computing and Communication Technologies (WCCCT)*, 2017.

[8] C. C. Wen *et al.*, "Analysis of Four Historical Ciphers Against Known Plaintext Frequency Statistical Attack," *Int. J. Integr. Eng.*, vol. 10, 2018.

[9] Y. Inan, "ANALYZING THE CLASSIC CAESAR METHOD CRYPTOGRAPHY," Oct. 2019.

[10] N. Bhandari, "Iterative Caesar cipher using grayscale image pixel values as keys," in *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2018, pp. 706–711, doi: 10.1109/PDGC.2018.8745973.

[11] S. M. Lokman, C. Chuah, N. H. A. Rahman, and I. R. A. Hamid, "A study of Caesar cipher and transposition cipher in Jawi messages," *Adv. Sci. Lett.*, vol. 24, pp. 1651–1655, 2018.

[12] S. B. Steever, "The Dravidian Language Family," in *The Cambridge Handbook of Linguistic Typology*, A. Y. Aikhenvald and R. M. W. E. Dixon, Eds. Cambridge University Press, 2017, pp. 887–910.

[13] D. G. G. Devi, "A Braille Transliteration on Tamil Vowels and Consonants Text Image," 2018.

[14] S. Chatterjee, H. Nath Saha, A. Kar, A. Banerjee, A. Mukherjee, and S. Syamal, "Generalised Differential Cryptanalysis Check for Block Ciphers," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2019, pp. 1137–1140, doi: 10.1109/IEMCON.2019.8936149.

[15] M. Mishra and V. H. Mankar, "A chaotic encryption algorithm: Robustness against brute-force attack," *Advances in Intelligent Systems and Computing*, pp. 169–179, 2012.

[16] Dave, K. T. International Journal of Innovations in Engineering and Technology (IJIET) Brute-

force Attack “Seeking but Distressing”, 2013.

[17] A. Al-Sabaawi, “Cryptanalysis of Vigenère Cipher: Methods Implementation Survey,” in *2021 International Conference on Intelligent Technologies (CONIT)*, 2020, pp. 1–4, doi: 10.1109/CSDE50874.2020.9411383.

[18] A. Al-Sabaawi, “Cryptanalysis of Classic Ciphers: Methods Implementation Survey,” in *2021 International Conference on Intelligent Technologies (CONIT)*, 2021, pp. 1–6, doi: 10.1109/CONIT51480.2021.9498530.

[19] A. Sweigart, *Invent Your Own Computer Games with Python, 4E*. No Starch Press, 2016.

[20] Vijayalakshmi, B., & Sasirekha, N. Lossless Text Compression for Unicode Tamil Document, 2018