Invention of Arbitrary Encryption System Using Applied Mathematical Tree Structure

K.Berlin, S.S.Dhenakaran, KR.Sivabalan

Abstract—Problem detection based on the communication system is considered as an important aspect of networking systems. Data transmission has played a big role in today’s world wide activities. But it is a big challenge to safe the data and escape from the hackers. The main objective of this research work is to provide strong cryptosystem for text data. To protect privacy of data plenty of research work has done using different cryptosystem methods. To enhance the security level an efficient crypto technique is designed for preserving privacy of plain text data. Construction of crypto system is relies upon the two separate technique named as Applied Mathematical Tree Structure (AMTS) and Matrix Mathematical (MM) model. The selection of an algorithm is entirely depends on the sizes of plain text. If the size of plain text counted as 2n level of order then it will processing with AMTS model, data other than the 2n level order works with MM model. Key has constructed from within the source file. Through the analyzed of security parameters we ensure that our proposed scheme increasing the level of security by using of two crypto systems.

Index Terms—Communication system, Applied Mathematical Tree Structure, Matrix Mathematical Model, Cryptosystem, Cipher-text only attack, Known-Plain Text attack.

1 INTRODUCTION

In the present scenario due to the fast growing of internet and electronic commerce, cryptography has become critical to business transactions and legal exchanges. The internet has made the transfer of data easy and perform vital role of our day to day activities. Computers allow us to store and transmit data with ease. Data storage takes up much less space than a filling cabinet and data can be transmitted almost instantaneously using the internet. The ease and speed of data transfer have led to the development of faster and bigger systems to handle large amount of data. Some of these data are private and contains information that is only to be read by certain individuals. Also critical information has transmitted over the internet. This information should be protected from unauthorized eyes. To achieve this goal, encryption is the technique used to preserve the data.

The major areas of cryptographic architecture are public key cryptography and symmetric key cryptography. The symmetric key based algorithms are called conventional cryptographic algorithms or secret key algorithms. They are implemented using two types of ciphers called block ciphers and stream ciphers. Stream ciphers can encrypt a plaintext as bit by bit, on the other hand block ciphers encrypt data as block wise. Using these cipher types, the data are separated into chunks; those chunks are encrypted and decrypted based on a specific key. Stream ciphers are used more dominantly than block ciphers, as the chunk is encrypted as a bit by bit basis. This process is much smaller and faster than encrypting large chunks or block of data. This method uses a secret key which is shared by both sender and receiver of the message.

The process of encryption and decryption has done by same key. The Asymmetric key model uses two keys to fulfill the encryption process, one for encryption and another one for decryption.

The methodology which has applied for hiding data is called as cryptography. It is the mechanism of attain security to save the secrets as non readable format. When the original message is encoded by using encryption scheme, the resulting data is notified as a cipher text and it is access only by those who know the encryption and decryption process of that corresponding scheme.

Cryptography is an effective and efficient method of assure data security. Commonly cryptography has applied on data transmission; it is now increasingly existence utilized for protecting information at same as considerably. The means of encryption is convert the original data into meaningless to save the ciphers unless the key file cannot be provided. Personal information or sensitive data like banking records, medical data and financial exchanges are made via public online transactions.

So as a security concern, safe of sensitive data is a tedious process. To overcome these kinds of risk cryptography can assure the security among the secrets. Cryptographic techniques permit authenticated persons to transmit and store data securely on unprotected networks so that the personal data can be known only by the authenticated receiver. Crypto mechanisms can carry out the encryption process on the secrets [14] and gives encrypted data in the form of meaningless to the persons who have not recognize of the key. Learning of the key is imperative for decryption.

Cryptographic techniques can use code substitution for encryption like matching each letter for appropriate numeric; otherwise use more complicated encryption methods to strengthen the secrets. The methods used for encryption and decryption may be explained publicly but the key is kept as secret [12]. Encryption is preferable good method for secure bulk of data while it can be transmitted across unsecure channels. Cryptosystem techniques can be used as a
symmetrical or asymmetrical. In public key cryptography, the person wants to pick up the encrypted data by publishes of own public key, which is used to data encryption. This is safe for data because the encrypted cipher text can be decrypted only by the private key which is created using the person's public key. The apt sender can be authenticated though this process. In symmetric same key used for encryption as well as decryption, so sometimes it suffers by key distribution problem. Same key that has to be sent from one sender to receiver and the question is how key is to be share securely, sometimes which can be impossible.

The cryptosystems are categorized as two broad concepts such as stream cryptosystem and block cryptosystem. Generally block ciphers cryptosystem much difficult to break then stream ciphers[15]. At the same time block cryptosystem needs more memory to implement, the requirement of memory space is based on the number characters used. Other than this, the special kind of cryptosystem is named as one-time pad[13]. It is mathematically unbreakable and it uses separate key for every message individually.

2 Encryption Methodologies for Preserving Text Data

Encryption with reduced sizes of cipher text was achieved by Mani Arora et al. [1] with the name of MDS Algorithm for Encryption. Objective of this scheme is to reduce the sizes of cipher text by using of pre chosen mathematical function. Dictionaries are plays a main role instead of key. Two different dictionaries are used as primary and secondary dictionary. In nature primary dictionary is static and secondary dictionary works dynamically. The primary dictionary has codes for words, numeral data and alpha numeric. Generally any digits can occupy 16 bits of memory, but the code can occupy 12 bits of memory for every digits. So automatically this approach reduces the sizes of cipher text. It reduces the redundancy and data transmission time.

Attribute based searchable encryption scheme was proposed using cipher text-policy attribute-based encryption technique [2]. For authorized searching data owners allowed to encrypt index keyword. Revocable Identity Based Encryption (RIBE) scheme was proposed by Rui Zhang et al. [3]. In this mechanism authors considers Key Dependent Message (KDM) security along with RIBE.

Hardik Gandhi et al. [4] implemented encryption scheme to encrypt text data named as research on enhancing public key cryptography by the use of MRGA with RSA and N-Prime RSA. In first phase magic rectangle is constructed in the order of 32x48. Magic rectangle having 1536 values totally. Magic rectangle divide into 12 quadrants and each having 128 characters. In second phase plain text is converted into numerals depends on its corresponding position of magic rectangle. Finally numerals are encrypted using RSA and N-Prime RSA algorithms.

Duong-Hieu Phan et al [5] designed public key encryption method with constant sizes of cipher text and secret keys to achieve security of chosen cipher attack. By using the same set of parameters authors provides the functionalities of revocation scheme as well as broadcast encryption simultaneously. Also implements inclusive-exclusive broadcast encryption scheme which has designed for the purpose of working as a both revocation scheme and broadcast encryption scheme at the same time. Akif Akgul et al [6] implemented encryption scheme for secure text data in the name of Text Encryption by Using One-Dimensional Chaos Generators and Nonlinear Equations. Three distinct chaos generators are applied and analyzed to process the encryption method. Chaos generators named as logistic map, pinchers map, sine-circle map. To increase the communication security non-linear equations are applied. Encryption was applied with the use of non-linear function,

\[ f(x,m) = \frac{m(3x^4 - x^2 + \sqrt{x})}{5} \] (1)

x refers ciphers that are generated by the chaos maps and m refers the message to be encrypted.

Encryption mechanism was implemented by Sreeja Rajesh at al [7] to transfer securely between devices. Method was named as A Secure and Efficient Lightweight Symmetric Encryption Scheme for Transfer of Text Files between Embedded IoT Devices. To enhance the level of security tiny symmetric encryption algorithm (NTSA) designed. While text encryption key confusions made dynamically for each round of process. An author assures that NTSA method is much more efficient and secure.

Polymorphic encryption key allocation scheme was designed by Geagan et al[11], key allocation scheme is considered as additional security element by the implementation of subset tree.

3 The Proposed Work

In this research work our experiments are carried out using Java NetBeans, the encryption algorithm, decryption algorithm, key generation process are implemented in Java NetBeans.

The whole work has divided into two phases based on the length of input data. If input data length has 2n it will work with phase 1. The data length has other than 2n it will process by phase 2 of research work.

\[ \sum_{i=0}^{n} Size_{input} = 2^n \quad n = 8, 16, 32 ... n \in N \] (2)

Where, i represent the number of input character.

\[ \text{Phase selection} = \begin{cases} 1 & \text{if } size_{input} = 2^n \\ 0 & \text{otherwise} \end{cases} \] (3)

First phase of our proposed algorithm divides the source data into two different files. The second half of data is considered as key to process the first half of input data [8]. The equation 2.3 is used to select the phase of encryption [9]. The dividend data length should be in equal. The first half data has converted into Complete Binary Tree (CBT) [10]. Output of CBT data has changed into reverse level order. The characters in reverse level order have turn into corresponding ASCII value. Now the key generation has done by second half of the data, it has converted into ASCII. ASCII values have shuffled in the wise of odd even. Finally the shuffled outputs are merging and it is considered as key.
The second phase has worked when the source data length other than the $2^n$. Here also the whole source file has divided into two halves. The second half of the data is used to create key as per the mechanism of phase1. The first half of the data has converted into its corresponding ASCII value. This ASCII values are arranged as 8×8 matrix. Then shuffle the matrix values in row and column wise. The shuffled outputs are transformed into UTF value. The second half of odd and even positioned data has shuffled. The shuffled data has changed into UTF-8. UTF-8 has converted into its extended ASCII. Finally this extended ASCII value is used as key.

**Phase1:** Works with $2^n$ range of input values.

**Phase2:** Works with other than $2^n$ range of input values.

---

Fig 2. Complete binary tree from phase I

Reverse level order of CBT is converted as,

$$s_1: u r f o t e s e b o t s e l \text{ etc...} \quad (4)$$

Now the RLO values are converted to ASCII and named as C1 as per the details mentioned in algorithms of phase1.

---

Fig 3. System architecture for Phase II

S: Divide two halves of equal length as $(s_3)$, $(s_4)$;
P: Convert ASCII;
P: Construct 8×8 matrix;
P: Shuffle (rows, columns) values vice versa;
C1: Generate UTF values of P;
K: Shuffle $s_4$ (odd, even);
K: Compute UTF-8;
K: Extended ASCII (UTF-8);
C: C1 + K

Now C refers that final cipher text. Now suppose source data $s_1$ has length of other than the $2^n$ let it execute phase2 algorithm.

8×8 matrix has done using the ASCII values of $s_1$,
Shuffle the values of rows and columns; result of this step is done in the order of

\[
m = \begin{pmatrix}
115 & 101 & 116 & 111 & 102 & 114 & 117 & 108 \\
101 & 115 & 116 & 111 & 98 & 101 & 102 & 111 \\
108 & 108 & 111 & 119 & 101 & 100 & 105 & 110 \\
105 & 111 & 110 & 115 & 111 & 114 & 111 & 116 \\
104 & 101 & 114 & 112 & 114 & 111 & 98 & 108 \\
101 & 109 & 115 & 111 & 108 & 118 & 105 & 110 \\
103 & 111 & 112 & 101 & 114 & 97 & 116 & 105 \\
\end{pmatrix}
\]

Now the shuffled matrix data are replaced by UTF-8 values. The values of UTF-8 has replaced by emoji with the help of its Unicode.

### 4 EXPERIMENTAL RESULTS

In this experiment various sizes of plain text data has tested and implemented. Results of proposed experiment are depicted in table1.

**TABLE 1**

<table>
<thead>
<tr>
<th>Plain text</th>
<th>Cipher text</th>
</tr>
</thead>
<tbody>
<tr>
<td>The time taken for the three stages (key generation, encryption and decryption processes) using the proposed algorithm increases whenever the size of the plain text is increased.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows that the time taken for execution of encryption process and key generation for various sizes of input files. Key has changed dynamically based on the input file. Size of encryption key is exactly equal to the plain text data. Absolutely for every input data key generated as much as strong. So it’s hard to execute and find secret by the hackers.

**TABLE 2**

<table>
<thead>
<tr>
<th>Plain text size (KB)</th>
<th>Time taken for Key Generation (sec)</th>
<th>Time taken for Encryption (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.38</td>
<td>2.7</td>
</tr>
<tr>
<td>100</td>
<td>4.45</td>
<td>12.96</td>
</tr>
<tr>
<td>300</td>
<td>11.64</td>
<td>26.32</td>
</tr>
<tr>
<td>500</td>
<td>21.45</td>
<td>70.61</td>
</tr>
</tbody>
</table>

5 **ANALYSIS OF THE PROPOSED ENCRYPTION ALGORITHM**

Proposed encryption method has analyzed to ensure the security level. The analysis was carried out by using various attacks to secure data against them.

5.1 **Known-plaintext attack**

The attack of known plaintext is more effective when the hackers try to hack ciphers. Known plaintext attack commonly used simple substitution cipher method to break the ciphers. In the sense of substitution cipher, our proposed algorithm changing the ciphers dynamically with the use of reverse level order method. For example,

Table 2 shows that the time taken for execution of encryption process and key generation for various sizes of input files. Key has changed dynamically based on the input file. Size of encryption key is exactly equal to the plain text data. Absolutely for every input data key generated as much as strong. So it’s hard to execute and find secret by the hackers.

**TABLE 3**

<table>
<thead>
<tr>
<th>Plaintext 1</th>
<th>Cipher 1</th>
<th>Plaintext 2</th>
<th>Cipher 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>u</td>
<td>e</td>
<td>l</td>
</tr>
<tr>
<td>e</td>
<td>r</td>
<td>f</td>
<td>t</td>
</tr>
<tr>
<td>t</td>
<td>h</td>
<td>f</td>
<td>r</td>
</tr>
<tr>
<td>o</td>
<td>r</td>
<td>o</td>
<td>m</td>
</tr>
<tr>
<td>f</td>
<td>t</td>
<td>r</td>
<td>f</td>
</tr>
<tr>
<td>r</td>
<td>e</td>
<td>t</td>
<td>f</td>
</tr>
<tr>
<td>u</td>
<td>s</td>
<td>l</td>
<td>e</td>
</tr>
<tr>
<td>l</td>
<td>a</td>
<td>e</td>
<td>s</td>
</tr>
<tr>
<td>e</td>
<td>b</td>
<td>s</td>
<td>e</td>
</tr>
</tbody>
</table>

In table 3 ‘s’ has replaced by ‘u’ and ‘e’ has replaced by ‘r’. On the other hand in table 4’s’ has replaced by ‘e’ and ‘e’ has replaced by ‘l’. This analysis shows that the substitution data has changed dynamically and proposed method fully secured against known-plaintext attack.

5.2 **Cipher-text only attack**

Cipher-text only attack uses the method to break the code named as frequency analysis. Frequency analysis is the study of group of letters in a cipher text.
Table 4 shows that the frequency of English alphabets and ciphertext occur in maximum different. In English alphabets letter ‘e’ has the maximum frequency of 12.7. Such that in cipher alphabets letter ‘o’ has the maximum frequency of 15.87. Hackers try to replace ‘o’ for ‘e’, but for ‘e’ the corresponding cipher is ‘e’. So based on the frequency analysis our proposed method has tested of cipher-text only attack.

ACKNOWLEDGMENT

This article has been with the financial support of RUSA – phase 2.0 grant sanctioned via Letter No F.24-51 / 2014-U, Policy (TNMulti-Gen), Dept. of Edn. Govt. of India, Dt. 09.10.2018.

REFERENCES


