

Linguistic Schemes Encoding Text Message

K. Sarath Kannan^{*1}, Dr. S. S. Dhenakaran²

^{*1}Department of Computer Science, Alagappa University, Karaikudi, Tamil Nadu, India

²Department of Computer Science, Alagappa University, Karaikudi, Tamil Nadu, India

ABSTRACT

Today websites and mobile applications are attacked by malicious, Ransom ware software designed to block information. It means sharing of websites and mobile apps information has low security. Cryptography is an art of providing security to messages before sharing them on insecure communication lines. The main aim of the paper is to propose a innovative encryption method adopting multilingualism approach where text message symbols are replaced by linguistic characters. An Unauthorized person cannot recognize the encrypted data while data transmission occurs. This method uses characters of four Indian languages Tamil, Hindi, Telugu, and Malayalam. This method encrypts plaintext and produces unintelligible linguistic codes. The outcome of encryption has a mixture of linguistic characters which cannot be understood by individual. Further, the initial experiments and results have shown a promising security level for proposed multilingualism technique. The proposed method is language-dependent approach providing new perspective with excellent potential for yielding secret codes.

Keywords: Encryption, Text Mixing, Linguistic Characters, Secret Message.

I. INTRODUCTION

Since the rise of the Internet, one of the most important factors of information technology and communication has been the security of information. Cryptography is the practice and study of hiding information. Modern cryptography intersects the disciplines of mathematics, computer science and electrical engineering. Cryptography has provided techniques for securing vital information and sharing them to others over insecure communication media. Though many different methods have been developed for securing messages, still there requires new principles to keep the contents of a message secret. Applications of cryptography include ATM cards, computer passwords and electrical commerce. Information security uses cryptography to transform usable information into a form from that sender alone

can understand and others cannot realize the secret codes. A generic model of encryption process is shown below.

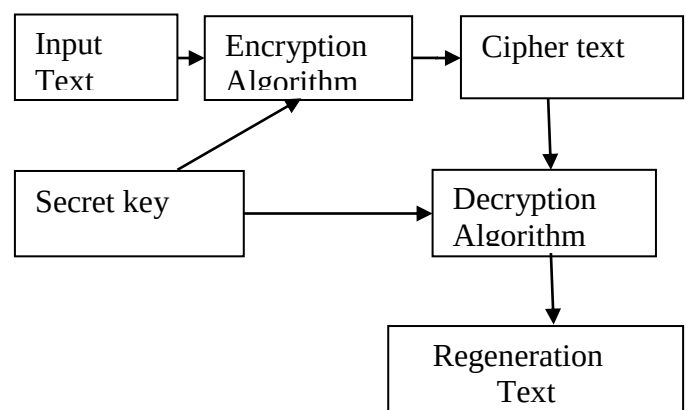


Figure 1: Generic Secret Key Encryption

It is obvious that a Multilingual person can talk in more than one language, either actively (through

speaking, writing, or signing) or passively (through listening, reading, or perceiving). More specifically, the terms bilingual and trilingual are used to describe comparable situations in which two or three languages are involved. A multilingual person is generally referred to as a polyglot. The proposed work considers text messages and converts them to multilingual symbols. These multilingual symbols are treated as the characters of secret code for the given input text message.

II. LITERATURE SURVEY

The Existing Encryption techniques are mainly dealt with European language like English, French, Germany etc. The Indian languages are used in the proposed work of Encryption Technique to generate secret message.

Arafat Awajan et al [3] implemented a coding system that provides a unique number for every character irrespective of the platform, program and language. The Existing Encryption techniques mainly deal with European language like English, French, German etc., The major disadvantage is that many characters are wrongly decoded in the decryption process.

B.Vijayalakshmi et al [4] discussed about the character encoding based on Tamil character. It is based upon the text characters. The limitation of this method is that, it is insufficient to represent all Tamil characters, store, transmit and retrieve the documents.

Md. Ahsamar Rahman et al [6] develops an algorithm for Bengali letters where the letters are encoded and the weights retain same order as the letters. And, compares the bit strings instead of weight strings which does not change the output. An ordering is maintained and Bengali letters are taken as basic input.

III. UNICODE MAPPING

The Unicode Standard is a character coding system designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world and creates the foundation of global software for recognition of characters. In addition, it supports classical and historical texts of many written languages. For example, The Tamil Unicode range is U+ 0B85 to U+0B95. The Unicode for the character 'அ' is 0B85; the Unicode for the character 'ஐ' is 0B95.

The Hindi Unicode range is U+0900 to U+09FF. The Unicode for the character 'अ' is 0900. The Unicode standard reflects the basic principle which emphasizes that each character code has a length of 16 bits. Unicode text is simple to parse and process and Unicode characters have well defined semantics. Hence, Unicode is chosen as the encoding scheme for the proposed work. After classification the characters are recognized and a mapping table is created in which the Unicode for the corresponding characters are mapped.

IV. PROPOSED METHOD

The proposed encryption technique is based on multilingual approach. It accepts input text and replace each character by multilingual characters which are displayed as the outcome of secret code. It is the requisite cipher text. Initially, the functional part collects all text characters in an array whose index locations are taken the weight of each text symbol. Similarly, the multilingual characters are collected in another array whose index value is referred for outputting linguistic character. For a given input text, the weight of each character is applied to function $f(x)=(2x+1)$. The functional value is again considered as a weight to look into linguistic table to

provide resultant cipher character. Say, A, the input character is stored in an array whose index = 0 ($x=0$), then the functional value of $f(x)=(2x+1)$, giving value 1 indicates the location of lingual array denoting linguistic character **அ**. This method uses four south Indian regional letters namely Tamil, Hindi, Malayalam and Telugu for collecting lingual characters. The functional process of the proposed work is depicted in Figure 2.

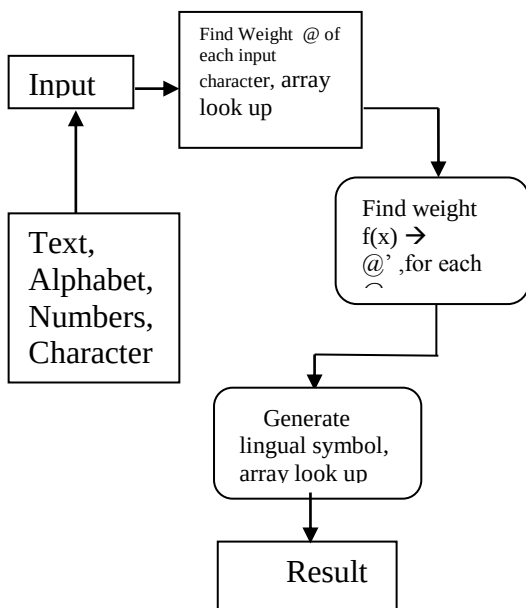


Figure 2: Functional Structure

Table1.Tamil – Unicode Symbol

S. No	Input value	Input text	Unicode	Cipher text
1	0	A	"\u0B85",	அ
2	1	B	"\u0B86",	ஆ
3	2	C	"\u0B87",	இ
4	3	D	"\u0B88",	ஈ
5	4	E	"\u0B89",	உ

Table 2.Hindi – Unicode Symbol

S. No	Input value	Input text	Unicode	Cipher text
-------	-------------	------------	---------	-------------

1	6	F	"\u0905",	अ
2	7	G	"\u0906",	आ
3	8	H	"\u0907",	इ
4	9	I	"\u0908",	ई
5	10	J	"\u0909",	उ

Table 3.Telugu - Unicode Symbol

S. No	Input value	Input text	Unicode	Cipher text
1	11	K	"\u0C05",	క
2	12	L	"\u0C06",	ల
3	13	M	"\u0C07",	ఇ
4	14	N	"\u0C08",	ఈ
5	15	O	"\u0C09",	ఉ

Table 4.Malayalam - Unicode Symbol

S. No	Input value	Input text	Unicode	Cipher text
1	16	P	"\u0D05",	അ
2	17	Q	"\u0D06",	ആ
3	18	R	"\u0D07",	ഇ
4	19	S	"\u0D08",	ഈ
5	20	T	"\u0D09",	ഉ

The above tables show a sample of few Tamil, Hindi, Telugu and Malayalam characters with Unicode.

