

Linguistic Schemes Encoding Text Message

K. Sarath Kannan¹, 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 multi lingual symbols. This multilingual symbol 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. France, 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+OB95. The Unicode for the character ' \mathfrak{B} ' is OB85; the Unicode for the character ' \mathfrak{B} ' is OB95.

The Hindi Unicode range is U+0905. The Unicode for the character ' \mathfrak{A} ' is 0905. The Unicode standard reflects the basic principle which emphasizes that each character code has a length of16 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 lingual 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 \mathfrak{B} . 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					
	Inpu				

S. N o	Inpu t value	Inpu t text	Unicode	Ciphe r text
1	0	А	"\u0B85",	அ
2	1	В	"\u0B86",	ஆ
3	2	С	"\u0B87",	Ø
4	3	D	"\u0B88",	١Ŧ
5	4	E	"\u0B89",	බ

Table 2.Hindi – Unicode Symbol

S.	Inpu	Innu		Cipho
Ν	t	Inpu t tort	Unicode	cipile r toyt
0	value	i lexi		i text

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

Table 3. Telugu - Unicode Symbol

S. N o	Inpu t valu e	Input text	Unicode	Ciphe r text
1	11	K	"\u0C05" ,	ອ
2	12	L	"\u0C06" ,	ື
3	13	М	"\u0C07" ,	શ
4	14	Ν	"\u0C08" ,	ఈ
5	15	0	"\u0C09" ,	ۍ

Table	4.Malay	alam -	Unicode	Symbol

S. N o	Inpu t value	Inpu t text	Unicode	Ciphe r text
1	16	Р	"\u0D05" ,	ത്ത
2	17	Q	"\u0D06" ,	ആ
3	18	R	"\u0D07" ,	ଗ
4	19	S	"\u0D08" ,	ഈ
5	20	Т	"\u0D09" ,	ତ୍ର

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

V. RESULTS AND DISCUSSIONS

Three examples are shown below letting different text messages are transformed into secret code. The secret code are mixture of lingual characters taken for processing. It shows that it is difficult to revert the secret code into text messages. When the processing includes symbols of more languages, it may be tedious to think of the original message. Also, the intruder must a multi linguistic to think of original contents which means the intruder must know all world languages. That is the challenge of the work in decoding the secret code.

Example:1

run: Input Text Is: This is Alagappa University Cyber Text Is: എറഡധഡഡക്ട്രസ്മക്തുക്തുക്തുണ്ട്ഡെട്രോഗേഷപണ്

Example 2:

Example 3:

Input Text Is:

This is Alagappa University.It is category 1st university in Tamil Nadu with A+ status.

Cyber Text Is:

ഥധഢപദഢൗഎന്വപരക്ഷാം പറക്ഷാം പാലാം വാലാം പാലാം പാലാം പാലാം വാ

ധപறപകധച

V. CONCLUSION

Multilingualism technique is based on Unicode characters in Tamil, Hindi, Malayalam and Telugu. Input of multilingualism technique accept all keyboard characters, numbers and special characters from Indian languages (Tamil, Hindi, Malayalam and Telugu) and as the result it is replaced with linguistic characters. The proposed method uses multilingualism for encryption. It is simple and efficient to use and is specially made for passing secret message. So that intruders cannot easily intrude the message.

VI. REFERENCES

- GanapathirajuM., BalakrishnanM., BalakrishnanN., and Reddy R., "Om: One tool for many (Indian) languages,"Journal of Zhejiang University Science, vol. 6A, no. 11,pp. 1348-1353,2005.
- R Seethalakshmi et.al., "Optical Character Recognition for Printed Tamil Text using Unicode", Journal of Zhejiang University Science, Vol. 6, No. 11, pp. 1297-1305, 2005.
- Arafat Awajan and Enas Abu Jrai, "Hybrid Techniques for Arabic Text Compression", Global Journal of Computer Science and Technology, Vol. 15, No. 1, pp. 23-27, 2015.
- B.Vijayalakshmi and N. Sasirekha," Lossless Text Compression For Unicode Tamil Documents" Department of Computer Science, Vidyasagar College of Arts and Science, India, Volume: 08, Issue: 02
- A. Bharath, S. Madhvanath, "Hidden Markov Models for Online Handwritten Tamil Word Recognition", 9th IEEE International Conference on Document Analysis and Recognition, 2007, Curitiba, Brazil, pp. 506-510.
- Md. Ahsanur Rahman, and Md. Abdus Sattar," A New Approach to Sort Unicode Bengali Text" 5th International Conference on Electrical and Computer Engineering ICECE 2008, 20-22 December 2008,
- SDivakaran, C.L. Biji, C.Anjali and Achuth Sankar S. Nair, "Malayalam Text Compression", International Journal of Information Systems and Engineering, Vol. 1, No. 1, pp. 7-11, 2013.
- Siva Jyothi Chandra, Ashlesha Pandhare and Mamatha Vani, "Multilingual Font Creation by Mapping Unicode to ASCII", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, No. 9, pp. 12-18, 2015.
- Ajantha Devi and S.Santhosh Baboo, "Embedded Optical Character Recognition on Tamil Text Image using Raspberry Pi", International Journal of Computer Science Trends and Technology, Vol. 2, No. 4, pp. 11-15, 2014.
- Siva Jyothi Chandra, Ashlesha Pandhare and Mamatha Vani, "Multilingual Font Creation by Mapping Unicode to ASCII", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, No. 9, pp. 12-18, 2015.