IJCTA, 9(25), 2016, pp. 475-483 ISSN: 0974-5572
© International Science Press

Security System Using Vernacular Languages

Rajbathi R.* and Diwakar R. Marur**

ABSTRACT

The principal goal of this paper is to design a security system which must be secure against unauthorised users. In general, the accumulation of data communication over digital computers increases rapidly in the recent years and this must be protected. In the present scenario, irrespective of the complex algorithmic designs, security systems are vulnerable to attacks by effective cryptanalysis. In order to achieve a secure transfer, a wide variety of encryption and decryption techniques are used but cryptanalysis has cracked it from time to time.In order to enhance the security system, vernacular languages are being used. The encryption and decryption techniques used in this security system is the multi-language encryption technique which are supported by Unicode.

Keywords: Unicode, Encryption, Decryption, Cryptanalysis, Security

1. INTRODUCTION

The growth of internet is rapid in the recent days and the wide spread availability of networks have led to the development of powerful and creative applications namely e-commerce, banking and communication of military messages. Privacy and security are two major problems imposed on e-commerce consumers and sites. There are many vulnerabilities in the e-commerce environment. Even in a simplified e-commerce environment scenario, a single user contacts a single web site, and then gives hiscredit card and address information for shipping a purchase, many potential securityvulnerabilities exist. The threat to loss of privacy, confidentiality, data abuse or misuse is about 42% [1]. In order to prevent the cryptanalytic attack cryptography comes into play.

The earliest use of cryptography was found in nonstandard hieroglyphs carved into monuments from the Old Kingdom of Egypt circa 1900 BCE. Later, Hebrew scholars made use of simple monoalphabetic substitution ciphers (such as the At bash cipher) beginning perhaps around 500 to 600 BCE. Cryptography has a long and complex history, it wasn't until the 19th century that it developedanything more than ad hoc approaches to either encryption or cryptanalysis.In World War II, mechanical and electromechanical ciphermachines were in wide use, where such machineswere impractical yet manual systems continued in use. Greatadvances were made in both cipher design and cryptanalysis [2].

Cryptography is widely used technique for protecting the communication of information over networks [3, 4]. Hence, the fundamental task of cryptography isnot only to protect the confidentiality of messages which are transmitted overpublic communication lines but also to resist suchcryptanalytic attacks which tend to evolve with the passage oftime [5]. It can be done by encoding or transmuting the data into an unreadable format. The original text is converted into anopaque equivalent text called cipher text and this process is called as encryption and the reverse is called as decryption. Encryption algorithms plays a major role in the information security system. Encryptionis the process that convertstheplain text to a cryptic text to secure it against data thieves.Decryption algorithm is the reverse of encryption algorithm. It is the process

* Department of Electronics and Communication Engineering, SRM University, Chennai, India

** Department of Electronics and Communication Engineering, SRM University, Chennai, India

of converting cryptic text to a plain text [6-10].Although, a widespread of techniques have beenemployed for encryption and decryption, the use of amultilingual approach for the same is not rampant.

Unicode concept is introduced so as to handle the Multilanguage text to achieve globalization. It overcomes restriction while using ASCII (American Standard Code for Information Interexchange) code. It enables the user to use more number of keys than ASCII permits. Unicode supports about 100 languages as of now [11, 12]. By using the help of the Unicode, a multilingual approach to cryptology can minimise the cryptanalysis. Enhancement in security canbe brought by localization of encryption technology.Linguistics plays a major in the security system using cryptography. The existing security system has a disadvantage that anyone can easily crack the code so vernacular languages were brought into cryptography for security system.Since, the combination of the vernacular letters are more, cracking of the code is the tedious process, so vernacular languages are preferred in this project.Multi-Language Encryption Technique (MULET) proposed by Praveen Kumaret al. is an encryption scheme which is designed to facilitate encryption/ decryption fora range of languages supported by Unicode. The authors have explained that thescheme is secure against brute-force attack only and have not conversed thesecurityagainst cryptanalytic attacks [13]. However, this scheme doesn't considered the attention of cryptanalysts.Anoop Kumar et al. indicated some of the flaws in the technique,no comprehensive cryptanalysis was presented [14].

The plain text is converted to cipher text by encryption technique and the encryption is done using Tamil Characters. Once the cipher text is obtained the Tamil characters are mapped with the Hindi characters. Decryption is the reversal of the encryption process.The Tamil language is preferred in order to make the code cracking impossible.Tamil is an ancient language like Sanskrit and not derived from any. It is the parent for any Dravidian languages (Malayalam, Telugu and Kannada).One of the ancient languages of the world still retaining its ancient flavours. It is a crispy language and a language with a very rich literature content. It is one of the few ancient world languages with explicit grammar. Tamil is a highly developed and dynamic language wherein prepositions may either be spelt separate or be spelt merged with nouns. The script is its own and not borrowed and modified, as with many other languages including English [15].

In the modern era, for network based systems many encryption techniques are evolved. They are of two types asymmetric key encryption techniques including RSA (Public Key Encryption) [16]& ECC (Elliptic Curve Cryptography) [17] and symmetric key encryption techniques including AES (Advanced Encryption Standard) [18] & DES (Data Encryption Standard) [19]. All the encryption techniques are being evolved for the English language. Since at the beginning only English is being used over the internet for the worldwide communication. The encryption techniques can be cracked by the third person and as soon as the plain text is captured it can be easily deciphered into English with the help of online translator. To overcome this problem a new technique named Tamil cipher is evolved. This method is created to enhance the security level of data. Tamilan cipher is a new encryption tool which is used for the encryption of regional language (Tamil). In English language there are only 26 characters available. Although it has been classified into two types such as vowels and non - vowels when considering for substitution techniques all the characters are being taken. Whereas in Tamil language there are about totally 247 character available. These 247 characters are further classified into four types namely*Uir, Mei, Uirmei and Aiudham*. In the existing system the plain text uses Tamil and Cipher key as Tamil Unicodeis used in numbers. This work proposes to make the plain text in Tamil and Cipher key in Tamil[20].

2. PROPOSED SYSTEM

2.1. Encryption

2.1.1. Plain text to Cipher text Conversion

Plain text refers to textual data in ASCII format. Plain text is the most portable format because it is supported by nearly every application on every machine. It is quite limited, however, because it cannot contain any formatting commands. Here the plain text used is the Tamil characters. Cipher text is encrypted text. Plaintext is what you have before encryption, and cipher text is the encrypted result. The term cipher is sometimes used as a synonym for cipher text, but it more properly means the method of encryption rather than the result. The cipher text which is used is the Tamil characters. Encryption is used to convert the plain text into cipher text. The multi-language encryption technique is used here for the Tamil plain text. Converting the obtained cipher text Tamil into Hindi to make the cipher text more complex for the security system. The Tamil characters are classified as vowels, consonants and Tamil signs. The Tamil Vowels are அ. ஆ, இ, ஈ, உ, ஊ, எ, ஏ ஐ ஒ ஓ. ஔ which has about 12 Characters. The Tamil Consonants are க, ங, ச, ஞ, ட. ண. த. ந. ப. ம, ய, ர, ல, வ. ழ. ள, ற, ன which has about 18 characters. The Tamil Signs are ◌ா, ◌ி, ◌ீ, ◌ு, ◌ூ, ெ◌, ே◌, ை◌, ெ◌ா, ே◌ா, ெ◌ள, ◌் . The dotted lines can be substituted by any of the letters from consonants. The combination of Tamil consonants and Tamil signs has about 216 characters. After Tamil character conversion it is mapped with Hindi character.

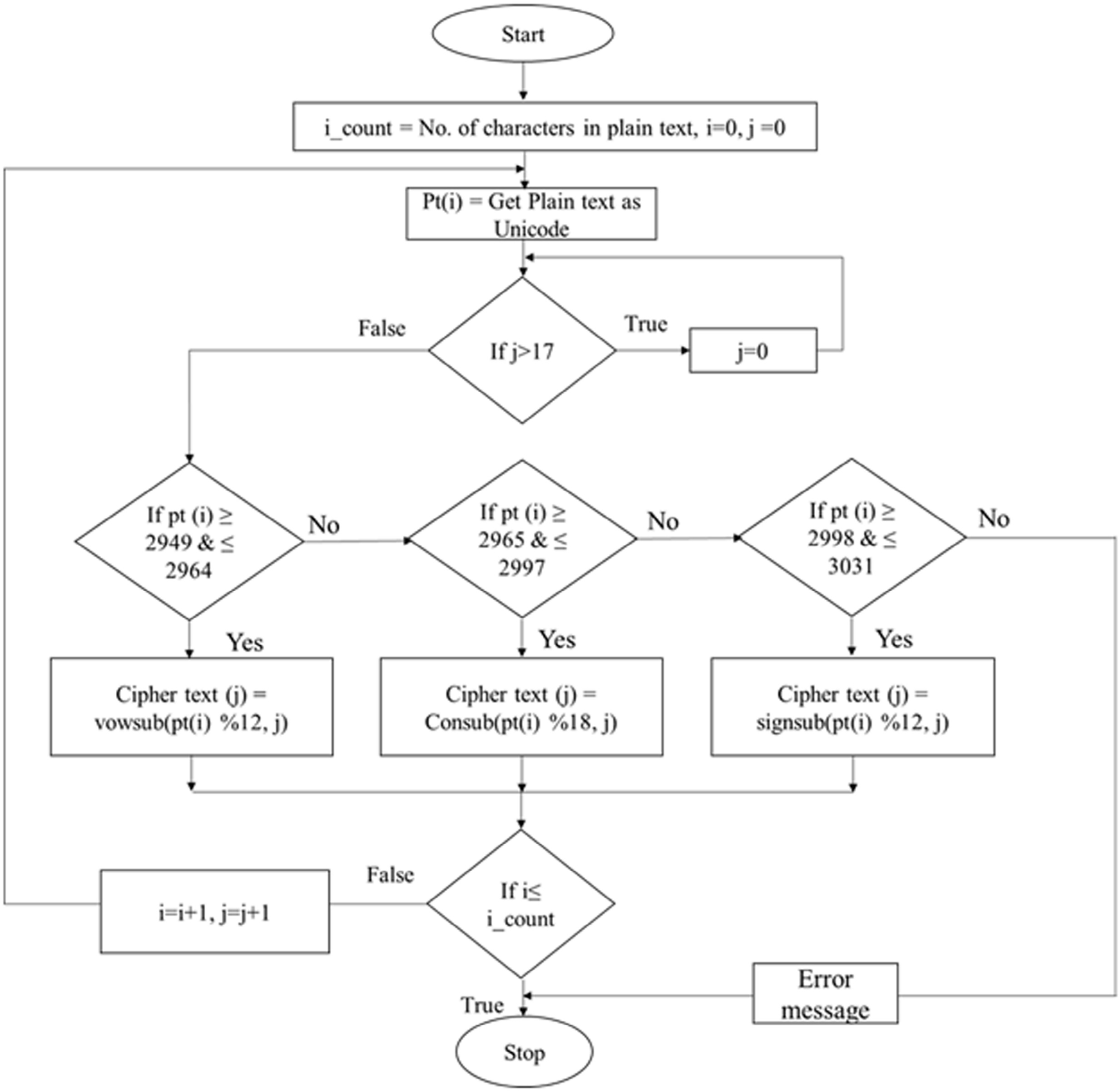

Start
i_count = No. of characters in plain text, i=0, j =0
Pt(i) = Get Plain text as Unicode
If j>17
True
j=0
False
If pt (i) ≥ 2949 & ≤ 2964
No
If pt (i) ≥ 2965 & ≤ 2997
No
If pt (i) ≥ 2998 & ≤ 3031
No
Yes
Cipher text (j) = vowsub(pt(i) %12, j)
Yes
Cipher text (j) = Consub(pt(i) %18, j)
Yes
Cipher text (j) = signsub(pt(i) %12, j)
False
i=i+1, j=j+1
If i≤ i_count
True
Error message
Stop

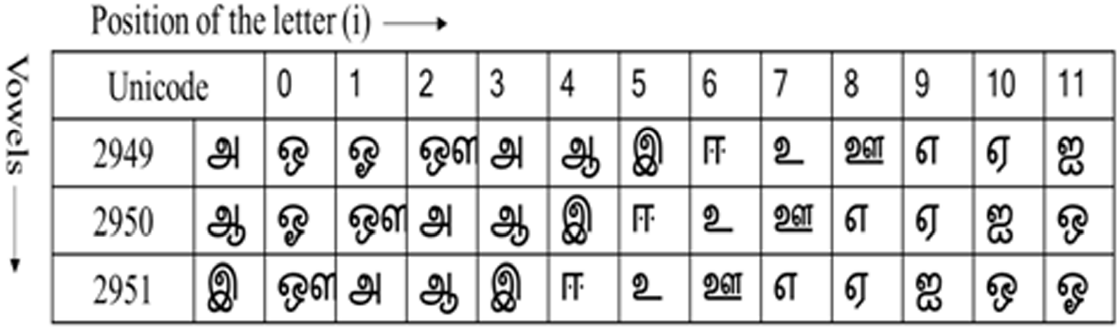
Figure 1: Flowchart for Encryption

Hindi is written in Devanagari script. It consists of 11 vowels and 33 consonants. Indian Government uses Hunterian transliteration as its official system of writing in Hindi.

Figure 1 indicates the flow of encryption process. In encryption, the plain text will be in Tamil characters. Once the plain text is obtained the Unicode of the particular text is retrieved. Once the Unicode is obtained, it will check for the type which falls under vowels, consonants, Tamil signs and punctuations or others. If the Unicode identified is vowels, the following procedure takes place. After the cipher text is obtained, it is mapped with the Hindi characters. The mapping of the Hindi (Devanagari) Characters is done in order to make the system more complex. For a single character in Tamil there will be four substitution characters in Hindi. This makes the code breaking near impossible.

If the Unicode is identified as "vowel" Tamil Character then the function Vowsub(pt(i), 12, j) is performed. pt(i) is the Unicode of the plain text, 12 is the total no. of letters in Vowels, j is the position of the letter in the word.The Unicode of Tamil vowels ranges from 2949–2964. Then the plain text is substituted with a new character from thesubstitution table to form the Cipher text. i.e., if "அ" is the plain text and assume it is in the position i = 0 and the unicode is 2949, if the substitution is performed then the cipher text which is obtained is "ஒ". The substitution takes place as follows. The unicode 2949%12 = 9. So the letter "அ" is replaced with the 9th letter of the vowel "ஒ" is replaced. And according to the position of the letter in a word the letters mapped are varied according to the given substitution table. The substitution table is given only for a few characters.

If the Unicode is identified as "Consonant" Tamil Character then the function consub(pt(i), 18, j) is performed. pt(i) is the Unicode of the plain text, 18 is the total no. of letters in Consonant, j is the position of the letter in the word. The Unicode of the Tamil consonants ranges from 2965 – 2997. Then the plain text is substituted with a new character from the substitution table to form the Cipher text. i.e., if "வ" is the plain text and assume it is in the first position i = 0, theUnicode is 2997, if the substitution is performed then the cipher text which is obtained is "ப".The substitution takes place as follows. The Unicode 2997%18

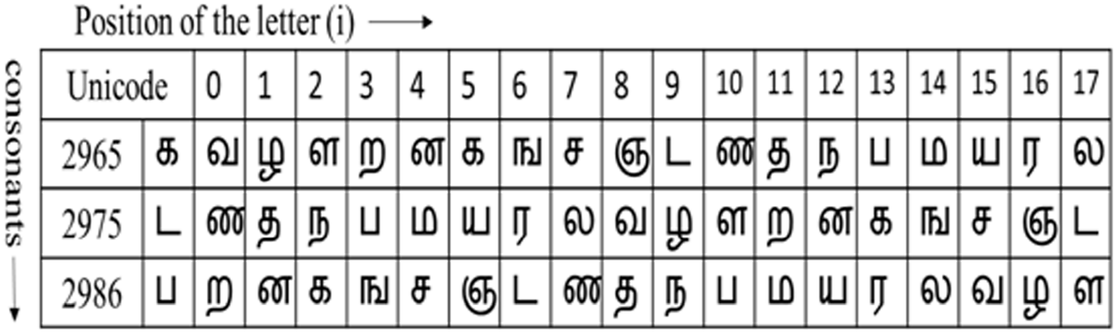


Position of the letter (i) →

Vowels ↓

Unicode		0	1	2	3	4	5	6	7	8	9	10	11
2949	அ	ஒ	ஓ	ஔ	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ
2950	ஆ	ஓ	ஔ	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ
2951	இ	ஔ	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ

Figure 2: Encryption Substitution table for "Tamil vowels"

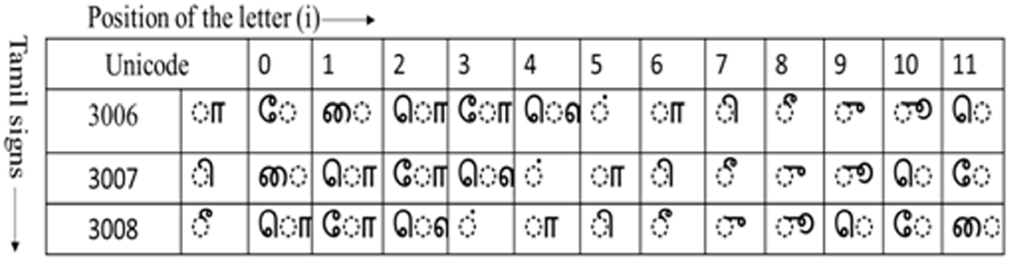


Position of the letter (i) →

consonants ↓

Unicode		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2965	க	வ	ழ	ள	ற	ன	க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ய	ர	ல
2975	ட	ண	த	ந	ப	ம	ய	ர	ல	வ	ழ	ள	ற	ன	க	ங	ச	ஞ	ட
2986	ப	ற	ன	க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ய	ர	ல	வ	ழ	ள

Figure 3: Encryption Substitution table for "Tamil consonants"



Position of the letter (i) →

Tamil signs ↓

Unicode		0	1	2	3	4	5	6	7	8	9	10	11
3006	◌ா	◌ே	◌ை	◌ொ	◌ோ	◌ௌ	◌்	◌ா	◌ி	◌ீ	◌ு	◌ூ	◌ெ
3007	◌ி	◌ை	◌ொ	◌ோ	◌ௌ	◌்	◌ா	◌ி	◌ீ	◌ு	◌ூ	◌ெ	◌ே
3008	◌ீ	◌ொ	◌ோ	◌ௌ	◌்	◌ா	◌ி	◌ீ	◌ு	◌ூ	◌ெ	◌ே	◌ை

Figure 4: Encryption Substitution table for “Tamil signs”

= 9. So the letter “வ” is replaced with the 9th letter of the consonant “ப ” is replaced. And according to the position of the letter in a word the letters mapped are varied according to the given substitution table.The substitution table is given only for a few characters.

If the Unicode is identified as “Tamil Sign” Tamil Character then the function signsub(pt(i), 12, j) is performed. pt(i) is the Unicode of the plain text, 12 is the total no. of letters in Tamilsign, j is the position of the letter in the word. The Unicode of the Tamil signs ranges from 2964 – 3031. Then the plain text is substituted with a new character from the substitution table to form the Cipher text. i.e., if “◌் ” is the plain text the Unicode is 3021, if the substitution is performed then the cipher text which is obtained is “◌ோ”. The substitution takes place as follows. The Unicode 3021%12 = 9. The dotted lines “◌் ” is made for substituting any of the letter from the consonants. So the letter “◌் ” is replaced with the 9th letter of the Tamil signs “◌ோ” is replaced. And according to the position of the letter in a word the letters mapped are varied according to the given substitution table.The substitution table is given only for a few characters.

2.1.2. Tamil to Hindi Conversion

The mapping of the Tamil characters with the Hindi characters is done in order to prevent the cracking of the code by the hackers. According to the position of the letter the respective Hindi character will be

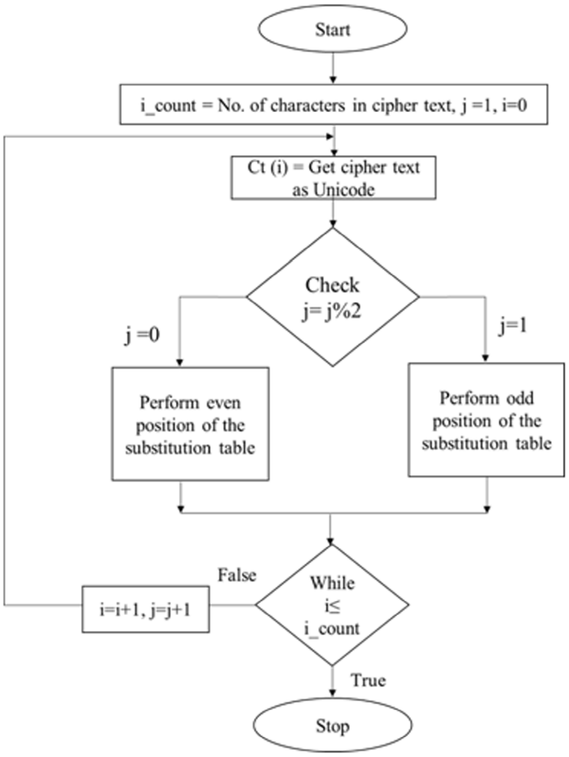
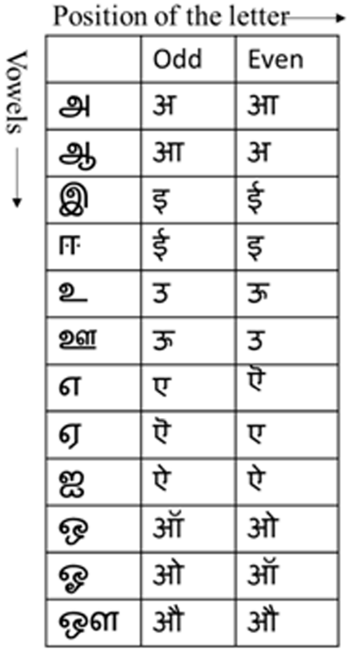

Start
i_count = No. of characters in cipher text, j =1, i=0
Ct (i) = Get cipher text as Unicode
Check j= j%2
j =0
j=1
Perform even position of the substitution table
Perform odd position of the substitution table
False
While i≤ i_count
i=i+1, j=j+1
True
Stop

Figure 5: Flowchart for “Tamil Vowels to Hindi Conversion”



Position of the letter→

Vowels↓

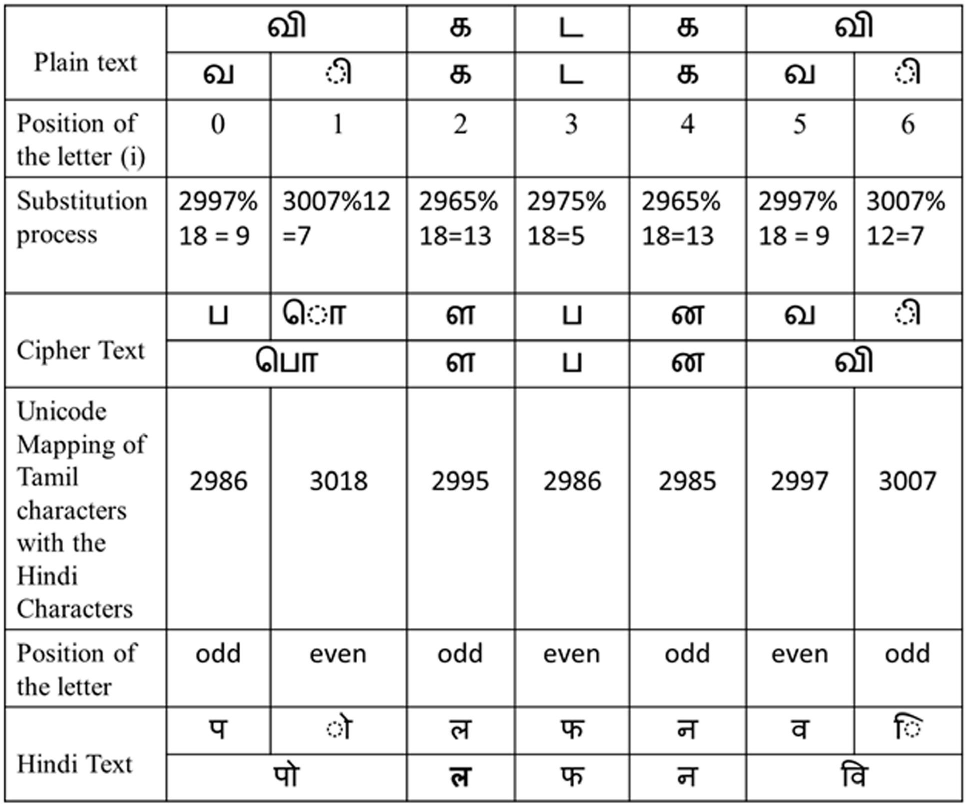
	Odd	Even
அ	अ	आ
ஆ	आ	अ
இ	इ	ई
ஈ	ई	इ
உ	उ	ऊ
ஊ	ऊ	उ
எ	ए	ऐ
ஏ	ऐ	ए
ஐ	ऐ	ऐ
ஒ	ऑ	ओ
ஓ	ओ	ऑ
ஔ	औ	औ

Figure 6: Substitution table for “Tamil Vowels to Hindi Conversion”

substituted as given in the substitution table. The substitution table mentioned here is only for Tamil Vowels. The Unicode for Devanagari Vowels ranges from 2309 – 2401. Assume ‘அ ” is in the odd position, Unicode is 2949%12=9 then it is replaced by the Hindi character “ अ ”.

For example the word “விவி க ட க ” is the plain text that is encrypted as “பொள பன வி” which is the cipher text. The substitution takes place as follows

Table 1
Encryption process

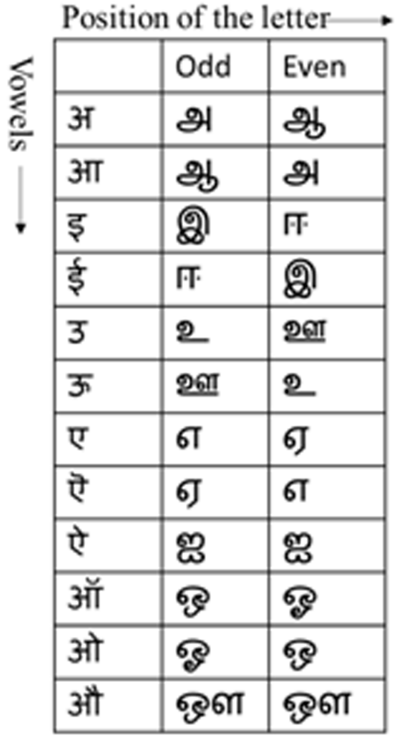


Plain text	வி		க	ட	க	வி	
	வ	ி	க	ட	க	வ	ி
Position of the letter (i)	0	1	2	3	4	5	6
Substitution process	2997% 18 = 9	3007%12 =7	2965% 18=13	2975% 18=5	2965% 18=13	2997% 18 = 9	3007% 12=7
Cipher Text	ப	ொ	ள	ப	ன	வ	ி
	பொ		ள	ப	ன	வி	
Unicode Mapping of Tamil characters with the Hindi Characters	2986	3018	2995	2986	2985	2997	3007
Position of the letter	odd	even	odd	even	odd	even	odd
Hindi Text	प	ो	ल	फ	न	व	ि
	पो		ल	फ	न	वि	

2.2. Decryption

2.2.1. Hindi to Tamil Conversion

Decryption is the reversal of encryption process. It is done in order to retrieve the original data. In decryption process, the Unicode of Hindi characters are obtained and are mapped with the Tamil characters. Then the Tamil Cipher text undergoes the substitution process then the Tamil plain text is obtained.



Position of the letter →

Vowels ↓

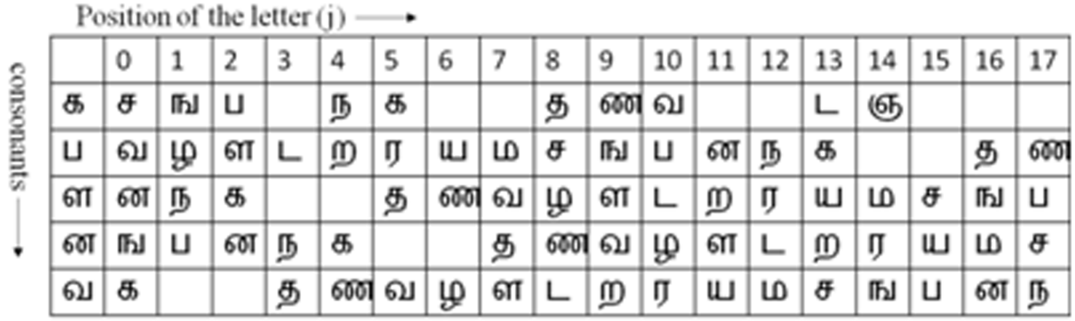
	Odd	Even
अ	அ	ஆ
आ	ஆ	அ
इ	இ	ஈ
ई	ஈ	இ
उ	உ	ஊ
ऊ	ஊ	உ
ए	எ	ஏ
ऐ	ஏ	எ
ऐ	ஐ	ஐ
ऑ	ஒ	ஓ
ओ	ஓ	ஒ
औ	ஔ	ஔ

Figure 7: Substitution table for "Hindi to Tamil Conversion"

Fig. 7 explains the substitution for" Hindi to Tamil conversion". In this conversion process the Unicode of the Hindi Cipher text is obtained. Then the Unicode of the Hindi characters are mapped with the Unicode of the Tamil characters. For the mapping to be done it must be checked whether the character is in the odd or even position. According to the position of the character the respective Tamil characters are substituted as in the substitution table.

2.2.2. Cipher text to Plain text conversion

In decryption, the cipher text of the Hindi characters is obtained; it is mapped with the Tamil characters. Then the cipher text of theTamil characters undergoes the decryption. Once the cipher text is obtained the

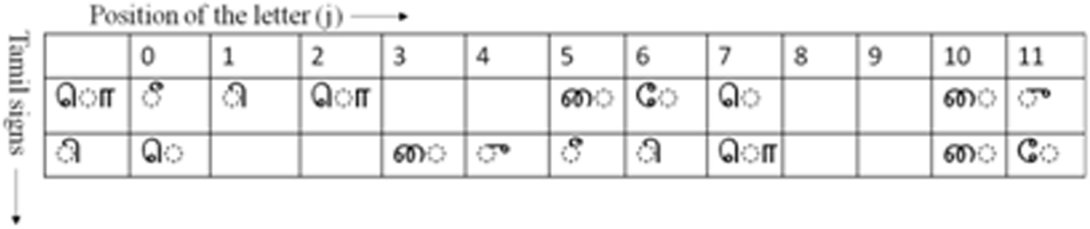


Position of the letter (j) →

consonants ↓

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
க	ச	ங	ப		ந	க			த	ண	வ			ட	ஞ			
ப	வ	ழ	ள	ட	ற	ர	ய	ம	ச	ங	ப	ன	ந	க			த	ண
ள	ன	ந	க			த	ண	வ	ழ	ள	ட	ற	ர	ய	ம	ச	ங	ப
ன	ங	ப	ன	ந	க			த	ண	வ	ழ	ள	ட	ற	ர	ய	ம	ச
வ	க			த	ண	வ	ழ	ள	ட	ற	ர	ய	ம	ச	ங	ப	ன	ந

Figure 8: Decryption Substitution table for Tamil Consonants



Position of the letter (j) →

Tamil signs ↓

	0	1	2	3	4	5	6	7	8	9	10	11
◌ொ	◌ீ	◌ி	◌ொ			◌ை	◌ே	◌ெ			◌ை	◌ா
◌ி	◌ெ			◌ை	◌ா	◌ீ	◌ி	◌ொ			◌ை	◌ே

Figure 9: Decryption Substitution table for Tamil Signs

Unicode of the particular text is retrieved. Once the Unicode is obtained, it will check for the type which falls under vowels, consonants, Tamil signs and punctuations or others. If the Unicode identified, then the following procedure takes place.

If the Unicode is identified as “Consonant” Tamil Character then the function consub(ct(i), 18, j) is performed. ct(i) is the Unicode of the cipher text, 18 is the total no. of letters in Consonant, j is the position of the letter in the word. The Unicode of the Tamil consonants ranges from 2965 – 2997. Then the cipher text is substituted with a new character from the substitution table to form the plain text. i.e., if “வ” is the cipher text and assume it is in the first position i=0, the Unicode is 2997, if the substitution is performed then the cipher text which is obtained is “க ”. The substitution table is given only for a few characters.

If the Unicode is identified as “Tamil Sign” Tamil Character then the function signsub(ct(i), 12, j) is performed. ct(i) is the Unicode of the cipher text, 12 is the total no. of letters in Tamilsign, j is the position of the letter in the word. The Unicode of the Tamil signs ranges from 2964 – 3031. Then the cipher text is substituted with a new character from the substitution table to form the plain text. i.e.,if ‘◌ி” is the cipher text the Unicode is 3007, if the substitution is performed then the plain text which is obtained is ◌ெ.The substitution table is given only for a few characters.

3. CONCLUSION

The aim of the work to make the security system more complex so that a high secured data transfer can be achieved. In order to enhance the security system, vernacular languages are being used. Vernacular languages are preferred because it cannot be easily cracked by any third person. The conversion between the characters of the vernacular languages makes the security system more complex. In this work the Unicode of the Tamil characters was retrieved and it is mapped to the Unicode of the Hindi characters. Using this the Tamil characters are converted to Hindi characters. Both the encryption and decryption are done using the Multilanguage encryption technique (MULET). The encrypted characters are merged with the conversion of Tamil characters to Hindi characters. The Hindi characters are passed on to the system after which the decryption takes place at the receiving end. The Hindi characters are converted to the Tamil characters and then the Tamil characters are decrypted and the original data was retrieved.

REFERENCES

[1] About e- commerce: http://www.uky.edu/~dsianita/390/390wk4.html

[2] History of Cryptography: https://en.wikipedia.org/wiki/History_of_cryptography.

[3] William Stallings, “Cryptography and Network Security Principles and Practices”, Prentice Hall, November 16, 2005.

[4] Ross J. Anderson, “Why Cryptosystems Fail”, Communications of the ACM,New York, USA, 1994, pp. 32-40.

[5] Brickell, E.F.; Odlyzko, A.M., Cryptanalysis: a survey of recent results, proc.of IEEE, issue 5 1998, pp. 578-593.

[6] William Stallings. Cryptography and NetworkSecurity Principles and Practices. Prentice Hall,November 16, 2005.

[7] Introduction to Cryptography. www.ipsec.com.Words: 725.

[8] John & Sons, Inc (Applied Cryptography). SecondEdition-New York, USA 1996.

[9] Frazier, R. E- (1998,1999) DataEncryption Techniques.

[10] D.B. Ojha, Ramveer Singh, Ajay Sharma, Awakash Mishra and Swati garg, “An Innovative Approach to Enhance the Security of Data Encryption Scheme. International Journal of Computer Theory and Engineering”, Vol.2, No.3, 2010.

[11] C.P.Ronald Reagan, S.Selvi, Dr.S.Prasanna Devi, Dr.V.Natarajan, “Enhancing DES Using Local Languages”, International Journal of Engineering Science and Innovative Technology (IJESIT)Volume 3, Issue 1, January 2014ISSN: 2319-5967.

[12] Unicode Character form http://www.unicode.org

[13] G. Praveen Kumar,Arjun Kumar Murmu,BiswasParajuli,Prasenjit Choudhury, “MULET: A Multilanguage Encryption Technique”, 2010 Seventh International Conference on Information Technology.

[14] Anoop Kumar Srivastava, Sanjeev Sharma, SantoshSahu, “MSMET: A Modified & Secure Multilanguage Encryption Technique”, International Journal on Computer Science and Engineering (IJCSE).

[15] Features of Tamil language: http://www.answers.com/Q/Is_Tamil_a_language

[16] R.L. Rivest, A. Shamir, and L. Adleman, “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems”, ACM communications 1978.

[17] Elliptic Curve Cryptography, Certicom Research, 2000

[18] AES Algorithm ,http://csrc.nist.gov/archive/aes/rijndael/wsdindex.html

[19] Walter Tuchman (1997). A brief history of the data encryption standar,ACM Press/Addison-Wesley Publishing Co. New York, NY,USA. 1997 pp. 275.

[20] P. Thamizhikkavi, Dr. S. Magesh, “Encryption Technology For Tamil Language (Tamilan Cipher)”, International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, April 2015.