# Enhancement in Encryption through Localization

**D.M.V.Sivakarthik**
Y8EM226, IV/IV B.Tech,
Department of ECM
KLUniversity, Guntur, India

**K. Ravi Kumar**
Asst. Proffessor
Department of ECM
KLUniversity, Guntur, India

**Abstract--The science of cryptology is a boon to the Internet era. In this age of universal electronic connectivity, of viruses and hackers, of electronic eavesdropping and electronic fraud, security matters the most. Ever since Caesar's time a wide variety of encryption techniques have been used but the cryptanalysis has simultaneously cracked these encryption techniques from time to time. Though complex encryption techniques have been employed in safeguarding data, the use of a multilingual approach is not prevalent. Unicode supports about 100 languages as of now. By using the help of Unicode a multilingual approach to cryptology can slow down the cryptanalysis multifold. In this paper, I aim to highlight the enhancement in security brought by localization of encryption and decryption. I also propose a novel algorithm that uses few cipher techniques already known to us transformed by the multilingual approach.**

**Keywords:** cryptology, multilingual approach, Unicode

## I . Introduction

The days that we considered computer as an advanced tool for only the elite but is neither useful nor affordable for common man are long gone. The innovations from the department of computer engineering in making the computer more useful have paved way to development of powerful and creative applications. The rapid growth of internet and widespread availability of networks have led to the massive usage of internet based applications[1].

Statistics reveal that on an average there are at least four globally connected / connectable (internet enabled) appliances in a house in case of developed countries and at least two globally connectable devices in case of developing and undeveloped countries. As a result the fact is evident that a huge amount of data which is data on the internet in server space, mail space and memory space offered by services and data transmitted over the internet for various purposes is at stake.

While data on internet is protected using advanced tools such as Intrusion Detection System (IDS), in most of the cases, it is during the transmission that compromising of data occurs. The attacks comprise of passive attacks such as release of message contents and more harmful active attacks such as masquerade and replay attacks.

## II . Intermediate Data Interception

The above attacks are possible only because the Intermediate Data Interception is possible. As you can see, in case of passive attacks the interception of data allows 'release of message contents' or 'traffic analysis' and in case of active attacks, the pattern recognition allows 'masquerade' and 'replay attacks'. Intermediate Data Interception is nothing but the interception of data at a point that is neither source nor intended destination of the data in the context.

This can cause chaos and/or damage because in this digitized world all the data transfer over the globe, and most of business transactions is online over the networks. Consider that the intelligence department has a suggestion for Police Headquarters which runs as "Delhi has been the recent target of terrorists, so it might not be targeted again anytime soon by the same as they will expect heavy security". Let us assume that the headquarters received the data and acted accordingly and let us suppose terrorists were able to intercept this data. This will cause serious damage as the terrorists might decide to target Delhi once again as they know police security might be less. Thus, Intermediate Data Interception must be prevented at all costs.

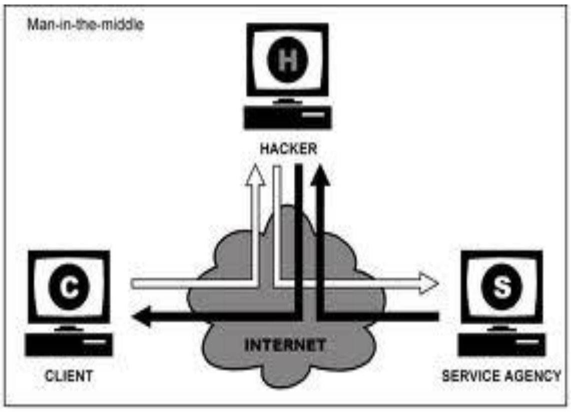


Figure 1 Intermediate Data Interception

The process of data interception to occur is possible because the middleman shares access to a router with sender & receiver of data which is being used for data transmission. As you know, over the network the data in general is too large to be sent in one piece, so it is split in to suitable sizes and sent in form of packets. The receiver receives these packets and positions them accordingly to get the data that is sent by transmitter.

For years encryption helped secure the data being transmitted over the network from intermediate Data Interception or middleman attacks. Encryption is nothing but the transformation of data using various algorithms that makes the data not readily intelligible. The data transforms back into sensible information only when the corresponding inverse algorithm is employed on the data. This inverse process is called decryption.

In spite of the advancement in encryption technology, Intermediate Data Interception are still possible the reasons of which are discussed below.

### A . Possible causes

It goes without saying that in spite of encryption if the middleman is able to get his hands on information regarding algorithm and the key for decryption, interception is invincible. Besides that, it is always possible to crack the encryption with cryptanalysis because,

- Encryption methods are known for years & are well analysed
- These methods tend to have limited language acceptance often English as it is the standard language for universal data transmission

### B . Cryptanalysis

Cryptanalysis uses different strategies for different encryption methods [3]. Let us see what cryptanalysis is based on in various methods.

a. Replacement strategy: To crack replacement strategy cryptanalysis uses probabilistic occurrence of characters and maps these with expected characters. These mapped characters are used to find the algorithm used by reverse engineering. It uses the concept that though replacement has taken place the probabilistic occurrence of a letter in a context in general doesn't change.

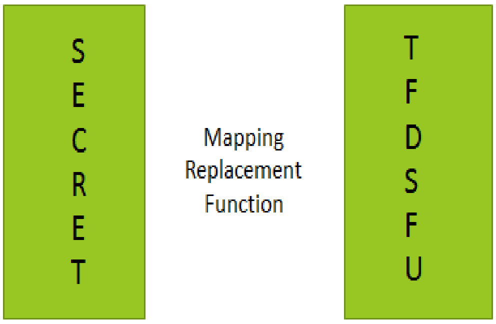


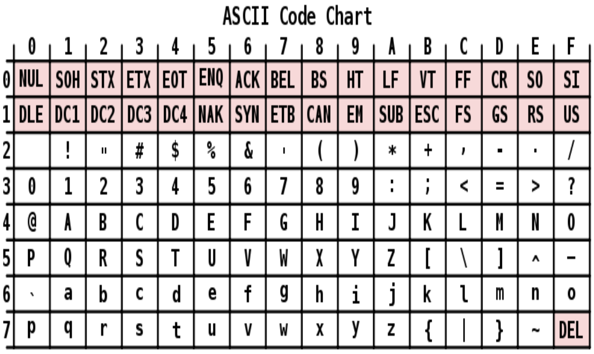
Figure 2: Replacement strategy example

b. Substitution algorithm: this technique exploits the fact that immediate repetition of characters will make identification of characters even easier and replaces the said characters with predefined character set based on a pre-decided strategy.

Cryptanalysis exploits the limited character set constraint on language and makes the distinction b/w replacement character set and substitution character set and uses it to crack encryption.[7]

## III . Localization approach

The main drawback of encryption is the limited character set availability of standard language- English. Only 256 characters are present in its character set.



ASCII Code Chart

|   | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | NUL | SOH | STX | ETX | EOT | ENQ | ACK | BEL | BS | HT | LF | VT | FF | CR | SO | SI |
| 1 | DLE | DC1 | DC2 | DC3 | DC4 | NAK | SYN | ETB | CAN | EM | SUB | ESC | FS | GS | RS | US |
| 2 |  | ! | " | # | $ | % | & | ' | ( | ) | * | + | , | - | . | / |
| 3 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| 4 | @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| 5 | P | Q | R | S | T | U | V | W | X | Y | Z | [ | \ | ] | ^ | _ |
| 6 | ` | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o |
| 7 | p | q | r | s | t | u | v | w | x | y | z | { | \| | } | ~ | DEL |

Figure 3 ASCII code chart

The localization approach is introduced to overcome this by making the use of various languages known to man such as Tamil, Urdu, and Hebrew etc. in encryption.[5]

### A . Unicode as a solution

This is possible using Unicode. Unicode is a computing industry standard for consistent encoding, representation and handling of text expressed in most of the worlds writing systems. ASCII is nothing but English character set in Unicode language library. With the release of Unicode 5.0, 93 languages have the corresponding characters digitized.[6]

### B . Character set Consideration

The localization approach requires that a character set is considered and used in encryption. It is to be noted that consideration of character set affects the efficiency of encryption.

A character set does not represent a language in its entirety. It is an extraction of characters accepted by the Unicode. So, it all comes down to characters that can be considered from scripts or writing systems. While characters in scripts can be in many languages, the characters in writing system would be confined. Because the discussion would be too vague when scripts are considered due to lack of proper classification writing systems are considered. The most popular writing system is latin.

In general, writing systems are classified into,

a. logographic
b. syllabic
c. alphabetic

Of the writing systems accepted by Unicode, the syllabic and alphabetic ones are easy to render relatively. The more random the mapping function is, the difficult it is to distinguish b/n substitution & replacement character set and aids to security. Though it’s a little difficult to render it is a good security measure to use logographic writing system.

A typical character set possible extracted from the Egyptician hieroglyphs writing systems is as follows.

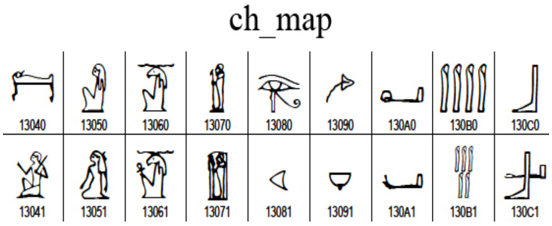


Figure 4 Replacement Character set

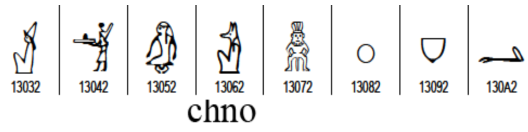


Figure 5 Substitution Character Set

### C . Applying Localization approach

Let us use a basic algorithm consisting of replacement and substitution methods to employ localization approach.

The text to be encrypted is read character by character and the Unicode value of each is obtained. This value is then divided by a mapping constant M. The remainder R so calculated is used as the index of mapping array ch_map and ch_map[R] is the corresponding cipher text. The quotient is stored in an array quo which is the key of encryption. Now on the cipher text Enc, replacement strategy is employed based on a pre-defined substitution array.

The decryption is as follows. The cipher text is scanned for characters in substitution array and reverse replacement strategy and mapping strategy is used. After obtaining the temporary decrypted text, the Unicode value of character of original message is calculated by adding R to product of M and Quo.

The main advantage or enhancement in security is that when bots try to decode the ciphered text without the key, the text would show alien languages unknown to the bot and hence the methods used on it would turn out to be futile. Further, the Unicode implementation is dependent on available storage, source code compatibility, and interoperability etc. The two major choices are Unicode Transformation Format (UTF) and Universal Character Set (UCS).

## IV . Conclusion

It is evident that the usage of localization approach enhances the security of encryption as we have discussed. Even the basic evaluation of strength of a normal algorithm like replacement and substitution as we have seen seems about a thousand times greater. Besides this the usability of Unicode is increasing in recent days i.e the implementation in everyday utilities like Office using fonts and in languages like perl making the use of various languages a reality. So localization can be seen as a rising trend in the security field.

## V . References

[1]. Che-Chern Lin "Study on internet usages, academic achievements, and the exploring capability of regional culture knowledge using internet" WSEAS Transactions on Information Science and Applications Volume 5 Issue 10, October 2008.
[2]. Ross J. Anderson, "Why Cryptosystems Fail", Communications of theACM, New York, USA, 1994, pp. 32-40.
[3]. Francois-Xavier Standaert, Gilles Piret, Jean-Jacques Quisquater,"Cryptanalysis of Block Ciphers: A Survey", UCL Crypto Group, 2003.
[4]. William C. Barker, "Recommendation for the Triple Data Encryption Algorithm (TDEA) Block Cipher", National Institute of Standards and Technology, NIST Special Publication 800-67, 2008.
[5]. Collins, R.W., "Software localization for Internet software, issues and methods", Software, IEEE, Florida, USA, 2002, pp. 74-80.
[6] Unicode Character form http://www.unicode.org
[7]. Multilingual Information Access 3-540-41933-0