

# THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

## A Survey on Telugu Text Steganography to Hide Encrypted Data

**B. Saritha**

M.Tech, Software Engineering & JNTU, School of Information Technology, India

**E. Jagadeeshwar Rao**

Assistant Professor , Software Engineering & JNTU, School of Information Technology, India

### **Abstract:**

*Telugu language is not similar to English, it does not contain characters as single alphabets as in English. Telugu is local language of Andhra Pradesh and it is widely used even in the defense department to transfer secret messages also. Now a days security of data is important, to provide security we have to hide or encrypt data using different techniques. In this paper we are using Steganography technique to hide data within the image. Introducing new encoding scheme is used to ensure more security.*

**Keywords:** Information Security; Text Steganography; Information Hiding; Text watermarking; Telugu Text Steganography

### **1. Introduction**

Telugu language is a Dravidian language used as local region language of Andhra Pradesh where we can see more number (about 75 million) of people speaking in Telugu only even to share secrets. Telugu characters consist of four types of alphabets where as English language only has one single type of alphabets i.e. A-Z. It consists of Vowels and Consonants, called as 'Achulu' and 'Hallulu' respectively. Gunintalu or Diacritics are formed with a combination of one consonant and one vowel. Compounds or Vattulu are formed with a combination of a consonant with another consonant.

Steganography is a way to hide text, image or a file with another text, image and file. It is used from the olden days. Encryption and steganography are different to one another. Encryption changes total data of a message to unreadable or not understandable form, whereas Steganography hides data of message i.e. it gives different appearance to the message itself. Steganography has several different techniques are present. Some of the techniques are: Digital watermarking, invisible ink, hologram etc.

### **2. Literature Survey**

The Existing systems use Java Unicode to read and store Telugu characters and encrypt data using specified algorithms such as AES, DES, Sha1, MD5 etc. By using this Unicode the size of the character may vary since Telugu characters are not plain alphabets like English. For single vowel or consonant Java Unicode size is 2-bytes which mean minimum size is 16-bits and maximum is 4-bytes. If Telugu character is simple vowel or consonant then the size of it is 2 bytes. If the character is a diacritic or component then it occupies 4 bytes. If the character is diacritic & component then it occupies 6 bytes. Anyone who knows Unicode can decode.

MS-Office supports Unicode feature by keying in the decimal code of given character along with 'Alt'- key. Similarly Java Unicode is available in many editing tools to use Telugu language so that it is easier to find out data even it is encrypted using Unicode.

To remove this problem a special encoding technique is introduced in this paper, where it deals with each and every letter of Telugu language up to four bytes only. Even it is not easy to recognize what data is sending for the intruder. Now a days usage of internet, mobile phones is increased which means digitization is increased. In this digitized world usage of emails, messaging is more than directly speaking. Hence we have to secure data transferred over the digital network. There are various encryption techniques are available for English language but not for Telugu.

The existing systems either support encryption or steganography over the Telugu language not the both. The main problem we are facing while transferring data is intrusion. There are several intrusion detection algorithms present which are usually detects intruder after data loss. To overcome this data loss we are introducing a gateway between sender and receiver. This gateway is like a trusted person where sender sends a copy to it and if receiver has any lost of data it must consult gateway for that missing data. The main difference between Encryption and Steganography is only the way they hide data. Steganography is also known as secret communicating science. The following figure shows the difference.



The proposed system is to design a new code for Telugu encryption so that it must has same size for any character of Telugu language. The size must also be acceptable i.e. which is less than 4 bytes. In this we have to implement cryptography which provides confidentiality, Integrity, Authentication for the text. Future work includes availability of text and checking whether there is any loss of data during transmission. Compared with the existing systems it saves memory up to 60%.

### 2.1. Various Methods

This section gives a brief view on different steganography algorithms. The advantages and disadvantages of the algorithms also included.

#### 2.1.1. Line Shifting Method

In Line shifting Method, Lines are shifted to some angle that is  $+\theta$  or  $-\theta$ . After the inclination the words in the inclined line represents data that to be transferred. But it is only useful for printed data otherwise we cannot use this method. For images or audio or video files transferring it is not useful. In this method the data can be retrieved using some character recognition programs.

#### 2.1.2. Word Shifting Method

In this method, words are shifted horizontally and the distance between words is changed to hide data. It is difficult to find out the data. It is acceptable and widely used where spacing between words does not matter. Sometimes it can also be used where distance between words matters since next line after paragraph ending is necessary. The main disadvantage is it consumes more time to retrieve information.

#### 2.1.3. Semantic Method

In this method, we use synonyms for certain words so that the data is hidden in those words. The main advantage of this method is it can't be retrieved using any character recognition programs but sometimes the meaning of data may be changed.

#### 2.1.4. Feature Coding

In this method, the data is elongated or shortened without affecting its meaning but it can be recognized by using retyping technique.

#### 2.1.5 Abbreviation

Hiding data in abbreviation forms is possible up to certain kilo bytes only. So we can't hide more data and sometimes the intruder may also has the possibility to find out since some abbreviations can easily recognized. For example: CS is for "Computer Science" and also for "Come Soon".

#### 2.1.6. Using HTML

In this method, by using HTML page we can easily hide data. HTML is not case sensitive which means `<BR>`, `<Br>`, `<br>` have same meaning. So we can hide data by changing source file of HTML page either in capital letters or in small letters. But it looks anonymous and can be understandable by end user if he sees source code.

#### 2.1.7. Pixel Masking

In this method the secret message is hidden in the image. Each pixel of image is altered with the encoded bits of secret message. For each pixel of R, G, B, the one bit value is added at the end which can't be recognized as a whole. But it can be recognized using optical character recognition programs if and only if the data to be hidden is large.

#### 2.1.8. Least Significant Bits coding

In this method, the least significant bits of cover image or cover message are replaced with secret message. Hence the total byte value is changed but not as a whole since each byte has a parity bit which is now changed to secret message.

#### 2.1.9. Video Steganography

Video is a set of image files, hiding information into a digital video format is known as video steganography. Here change in any one of the set of image cannot be recognized to human eye. Hence using this method we can hide image better than using image steganography to hide another image.

## 2.2. Related Work

From the survey that was conducted on different existing systems techniques the new project “Telugu Text Steganography to Hide Encrypted data” consists of following steps at the sender side:

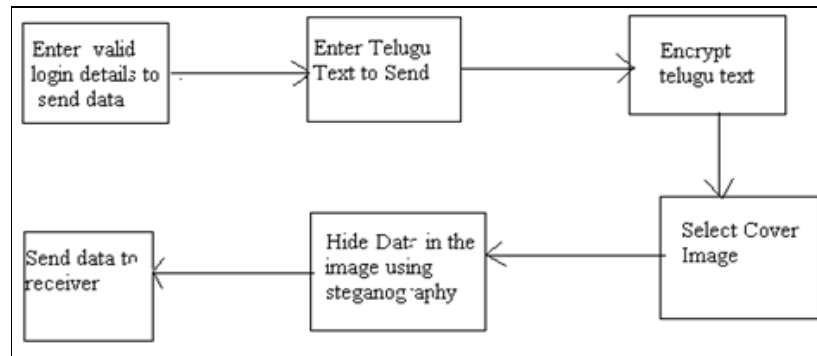


Figure 1: Steps at sender side

### 2.2.1. Enter valid Login Details

At this first step sender must has to login with login credentials to send data to the receiver. The data to be sent is entered by the sender. If the login credentials are not valid then sender could not send data. This is not a web based application hence first the user has to be registered online then he can login into this application (using java applets).

### 2.2.2. Send Data

Send data by entering data using keyboard and this application is mainly for Telugu text so sender has to enter Telugu text only.

### 2.2.3. Encrypt data

In this proposed system Encryption is done using our newly defined encoding scheme of 16-bits for each character of the entered Telugu text. Each character can be either a single vowel or consonant and can be a combination of vowel or consonant with diacritic or compound. Our new encoding technique has three regions namely R1, R2, R3. R1 is allocated for alphabet (either vowel or consonant), R2 is allocated for diacritic and R3 is allocated for compound. Size of each region is as follows: R1 is 6-bits, R2 is 4-bits and R3 is 6-bits, totally 16-bits.

To encode we have to assign values to each letter of Telugu language which ranges from 0-52 and 1-15 for vowels and diacritics, 16-48 for consonants and compounds. Thus each character occupies only 16-bits (2 bytes) as constant size as encoded character.

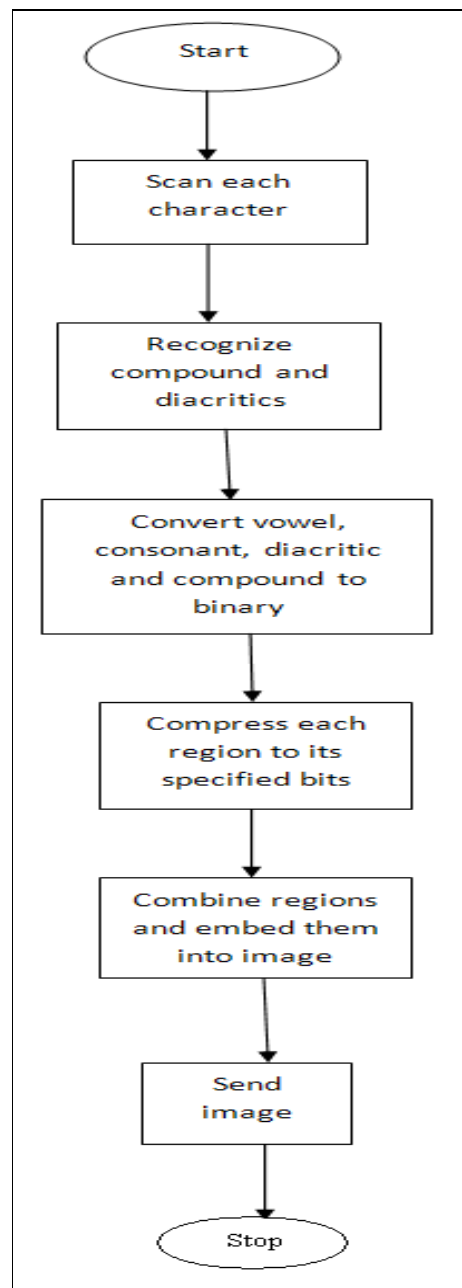


Figure 2: Flowchart of encryption

#### 2.2.4. Select image & hide data

Select image one of the images available in the directory folder to hide the encoded text into that image using Image Steganography. In this we are using pixel masking technique since we are sending data which is reduced its size up to 60% less than it is for Java Unicode, so pixel masking is efficient for this transfer.

#### 2.2.5. Send to Receiver

To send data to a receiver click send button after successful completion of Image Steganography process. After sending data the receiver also has the same steps which are in reverse to get the Telugu text.

To remove the loss of packets while sending data, the sender sends data to the gateway which is middle way between sender and receiver. If the receiver has noticed any loss of data then it requests for data from the gateway.

### **3. Conclusion & Future Work**

Even there are many ways to provide security to send data, combining cryptography and encryption gives more security than using them individually. Increase digitization results need for more security to secure data from intruders. Future work includes saving of particular packet loss and also determining that particular packet number and data.

**4. Acknowledgment**

This work is supported and made under the esteemed guidance of Mr. E Jagadeeshwar Rao, who is a lecturer and head of the department of Software Engineering at School of IT, Jawaharlal Nehru Technological University Hyderabad.

**5. References**

1. A NEW APPROACH TO TELUGU TEXT STEGANOGRAPHY , Dr. Ramineni Siva Ram Prasad,Acharya Nagarjuna University,Andhra Pradesh, INDIA.Kalavathi Alla,Vasireddy Venkatadri Institute of Technology.
2. A NEW ENCODING SCHEME OF TELUGU TEXT FOR INFORMATION HIDINGLAKSHMEESWARI G., RAJYA LAKSHMI D. AND SRINIVAS Y.GITAM Institute of Technology, GITAM University, Andhra Pradesh, India.
3. A NEW APPROACH TOTELUGU TEXT STEGANOGRAPHYBY SHIFTING INHERENT VOWELSSRAVANI ALAMETI M.TECH (DECS)SAKE POTHALAI AH M.E (ECE), Asst Professor of ECE, Dr.K ASHOK BABU Professor & HOD of (ECE), SRI INDU COLLEGE OF ENGG&TECHNOLOGY, Hyderabad, Andhra Pradesh, India