International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.6, June 2021 E-ISSN: 2347-2693

# A Novel Hybrid Symmetric Key Encryption Algorithm for Telugu Script

**T. Madhavi Kumari[1*], A. Vinaya Babu[2]**

[1]Associate Professor, Dept. of ECE, JNTUH College of Engineering, Hyderabad, Telangana, India
[2]Professor of CSE& Dean Academics, Stanley College of Engineering, Hyderabad, Telangana, India

[*]*Corresponding Author:thoomati@gmail.com, Tel.: +91 8008103823*

DOI: https://doi.org/10.26438/ijcse/v9i6.9196 | Available online at: www.ijcseonline.org

Received: 05/Jun/2021, Accepted: 15/Jun/2021, Published: 30/Jun/2021

***Abstract*—**Technology is the key to innovation in all aspects of this modern age. In any technology, data becomes the most important asset to protect. Many encryption algorithms are widely available and used in information security. Encryption can provide secure information across the platform. Encrypting the message using natural languages reduces encryption time and improves performance. Telugu is the oldest Dravidian language spoken in South India. The message to be encoded is translated into Telugu, after which this translated text is converted into a randomly generated combination of 2-bit English alphabets. This is a hybrid algorithm because the intermediate node is encrypted using the standard advanced encryption algorithm to improve the privacy of the text. Because Telugu encryption method uses three phases, namely translation, mapping, and encryption, this makes the data much more secure than existing algorithms like Blowfish and Data Encryption Standard (DES), and the person trying to decrypt must have knowledge of the Telugu language, as well as mapping details to see the original data and This algorithm shows a stronger avalanche effect of 96.5%, which is greater than Blowfish and DES. Evaluation of the proposed algorithm shows that it runs faster and has relatively less encryption time, less memory requirements.

***Keywords*—**Cryptography, Security, Telugu language, Encryption, AES

## I. INTRODUCTION

In the modern era, all tasks have been computerized and there is a lot of data transfer. It is very important to protect this data from unauthorized persons. Any data in someone else's hands is a threat, and people will not use these modern technologies if they feel insecure, which is why the crypto domain is a valuable domain and is the foundation of many modern applications. Cryptography is a means of protecting information through the use of codes [1]. Subsequently, this protected information can only be read by those for whom the information is intended to be read and processed. Encryption techniques protect sensitive data, such as credit card numbers, by encrypting and turning the information into unreadable cipher text. Later, this encrypted data can be decrypted or only read with a secret key [2,3].

The two main cryptographic techniques based on the number of keys used are the symmetric key and the asymmetric key. Symmetric key cryptography involves the use of a secret key in conjunction with encryption and decryption algorithms that help protect the content of the message [4]. Asymmetric key cryptography is also known as public key cryptography because it involves the use of a public key in conjunction with a secret key. Previously, the most common symmetric key encryption algorithm is the Data Encryption Standard (DES), which uses a 56-bit key and encrypts 64-bit data. Later, since DES was broken down by a cryptanalyst, symmetric block encryption algorithms such as Blowfish and Advanced Encryption Standard (AES) were developed. AES is considered efficient and reliable because it uses a 128, 192, or 256-bit key. Each algorithm has its drawbacks when it comes to protecting confidential data [5,6].

For more efficient encryption, you can use natural language text. Here we use the Indian language cart as it is 56 characters long and it is very difficult to break the message [7]. It is used in conjunction with AES to make the encryption process more secure. This robust process consists of three stages. The first stage is TRANSLATION, where the plain English text is changed to Telugu. This translation is done using the Google Translator API. This API translates the message to English in Telugu. The Telugu language is 56 characters long, and a person who wants to cryptanalyze data must be well versed in Telugu. The second phase is MAPPING, where each Telugu alphabet is mapped to two 2-bit combinations of English alphabets, and the result of this phase is called intermediate encryption. The final step is ENCRYPTION, which is to encrypt the intermediate encryption using an existing encryption algorithm (AES). The data or messages go through all these stages, so the data becomes more secure. In this algorithm, because the message is translated into the Telugu language, which has a higher number of alphabets, this makes the algorithm much more efficient than existing algorithms. Telugu Encryption provides a fast and secure encryption method that can be used by the banking sector and the government to transmit data in a much more secure way [8].

## II. RELATED WORK

Recent encoding algorithms play an important role in guaranteeing the protection of data technology and communications systems. It offers confidentiality, authentication, integrity and non-repudiation. Safiah et al. [9] proposed NASE (Novel Algorithm in Symmetric Encryption) as new formula in rhombohedra encoding that supports Feistel. NASE involves generating a random Block variety size, iterations and completely different keys for every block. The planned formula works for quite one language (for example, English, Arabic, etc.) and is feasible to use double or triple encoding with completely different keys to further improve the security. NASE is extremely quick and simply applied to completely different applications.

Gupta et al. [10] introduced a replacement cryptography technique that uses multiple set of language characters. They have worked alone at English and Hindi languages. In this paper Vidhya and Paul [11] proposed an encryption method using an Indian local language, Malayalam. The proposed method consists of custom Unicode based technique with embedding based on indexing, i.e. the original message is encoded to Malayalam text with custom UNICODE values generated for the Malayalam text. The proposed method is more precise in the encoding process and in the decoding process. The method achieved a precision rate of .95 and decoding rate of .81. Khairullah [12] presented a simple and novel approach for steganography through transliteration. A phonetic keyboard layout is very popular for writing languages having non-roman alphabets. The results show that the capacity of the method is 1.2%, which is adequate for a text steganography system with very low risk of machine detection. This method can be easily adapted and applied for any other language having non-roman alphabet.

Taha et al. [13] proposed an algorithm for information hiding using Arabic text. The new algorithm improves the length of the secret message that can be embedded in an Arabic text document without affecting its quality as much as possible. The proposed algorithm utilizes different characteristics and properties of Arabic language. The proposed algorithm was tested for different length stego-text messages. It provides superiority in achieving high capacity hiding ratio in comparison with the most related Kashida-based techniques and spaces-based techniques.

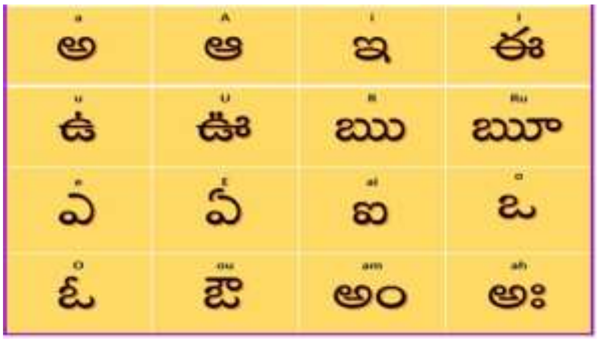
In this paper Hamzah et al. [14] proposed a framework that uses Arabic calligraphy to hide information. Roy and Venkateswaran [15] present a text based steganography technique based on the Vedic Numeric Code. Frequency of the letters in English alphabet in conjunction with Vedic Numeric Code is used for the steganography technique. No separate importance is given for vowels and consonants. Vijaya Bharati and Jyothi Prasad [16] deals with a practical scheme for encoding an Indian language, telugu. The proposed technique uses the Telugu Text and their attributes to hide the secret message. The Cover Message is essentially a collection of key combinations stored in the form of rows and columns. These combinations are generating by encryption of the saved Text documents. The attributes combinations used in the Text are used to generate a Cover Message. Changder et al. [17] presents a new linguistic approach through Indian Languages by considering the flexible grammar structure of Indian Languages. The proposed method exhibits satisfactory result on some Indian Languages like Bengali.

Even though many natural languages are used in the literature for the protection of information from being read by malicious third parties they still lag in protecting the text completely. Our proposed model is considered to be more robust than the other existing algorithms since it uses a language that supports more letters and uses AES in addition to improve the security.

## III. METHODOLOGY

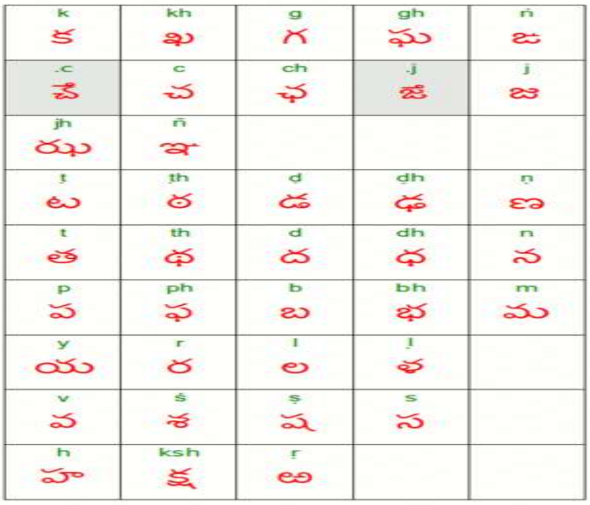
In the proposed system we propose a new hybrid symmetric key encryption algorithm which is more secure and shows strong avalanche effect (small change in the plaintext bit produces considerable change in the cipher text bits) by having a hard brute force. This Encryption algorithm makes use of Telugu language for encryption. These algorithms only generate an intermediate cipher. This is then given to AES algorithms to make it more secure. Even if cryptanalysts find the intermediate cipher he then need to handle AES algorithms to find out the plain text [18, 19].

Telugu script an abugida from the Brahmic family of scripts, is used to write the Telugu language, a Dravidian language spoken in the Indian states of Andhra Pradesh and Telangana as well as several other neighbouring states. The Telugu script is also widely used for writing Sanskrit texts and to some extent the Gondi language. It gained prominence during the Eastern Chalukyas also known as VengiChalukya era. It shares extensive similarities with the Kannada script, as it has evolved from Kadamba and Bhattiprolu scripts of the Brahmi family. In 2008, Telugu language was given the status of Classical Languages of India, this status owes to its rich history and heritage [20, 21]. Telugu content has 18 vowels and 36 consonants, of which 13 vowels and 35 consonants are in like way use and they are shown in Figure 1 and Figure 2.



| | | | |
|---|---|---|---|
| a<br>అ | A<br>ఆ | i<br>ఇ | I<br>ఈ |
| u<br>ఉ | U<br>ఊ | R<br>ఋ | Ru<br>ౠ |
| e<br>ఎ | E<br>ఏ | ai<br>ఐ | o<br>ఒ |
| O<br>ఓ | au<br>ఔ | am<br>అం | ah<br>అః |

**Fig. 1:** Vowels of Telugu Language



| | | | | |
|---|---|---|---|---|
| k<br>క | kh<br>ఖ | g<br>గ | gh<br>ఘ | ṅ<br>ఙ |
| .c<br>ౘ | c<br>చ | ch<br>ఛ | .j<br>ౙ | j<br>జ |
| jh<br>ఝ | ñ<br>ఞ | | | |
| ṭ<br>ట | ṭh<br>ఠ | ḍ<br>డ | ḍh<br>ఢ | ṇ<br>ణ |
| t<br>త | th<br>థ | d<br>ద | dh<br>ధ | n<br>న |
| p<br>ప | ph<br>ఫ | b<br>బ | bh<br>భ | m<br>మ |
| y<br>య | r<br>ర | l<br>ల | ḷ<br>ళ | |
| v<br>వ | ś<br>శ | ṣ<br>ష | s<br>స | |
| h<br>హ | ksh<br>క్ష | ṟ<br>ఱ | | |

**Fig. 2**. Consonants of Telugu Language

The plain text undergoes three phases in the proposed system the translation phase, the mapping phase and the encryption phase. The plain text in English language is first translated to Telugu language using an language translation API. Then the translated text is produced. This translated Telugu text enters the next phase. In the next phase each letter of the Telugu text is mapped with two English alphabets which is generated randomly, the result of this phase is called intermediate cipher. Now the intermediate cipher must be encrypted this is done using AES Encryption algorithm. To decrypt the encrypted text also consists of the same three phases but in reverse order. First the encrypted text (AES) is decrypted which gives the intermediate cipher. The intermediate cipher is mapped with 2-bit key which gives the translated text (text in Telugu). Now the out from the previous phase is translated to English to get the original message. AES uses longer keys, such as 128, 192 and 256 bits for encryption. Therefore, it makes the AES algorithm more robust against piracy. For 128 bits, about 2128 interruption attempts are required. This makes hacking very difficult, as it is a very secure protocol.

### 3.1 Translation

The plain text is translated from English language to Telugulanguage, this translation is done with the help of Google translator API. Numbers cannot be translated with this API, hence the numbers are first converted to words and then it is translated to Telugu. A language translator translates a text written in one language to other language. It is a very helpful tool allows people to understand text written in unknown language. Google translator is very famous and the best translator available, it is a lot easier to use and also free of cost.

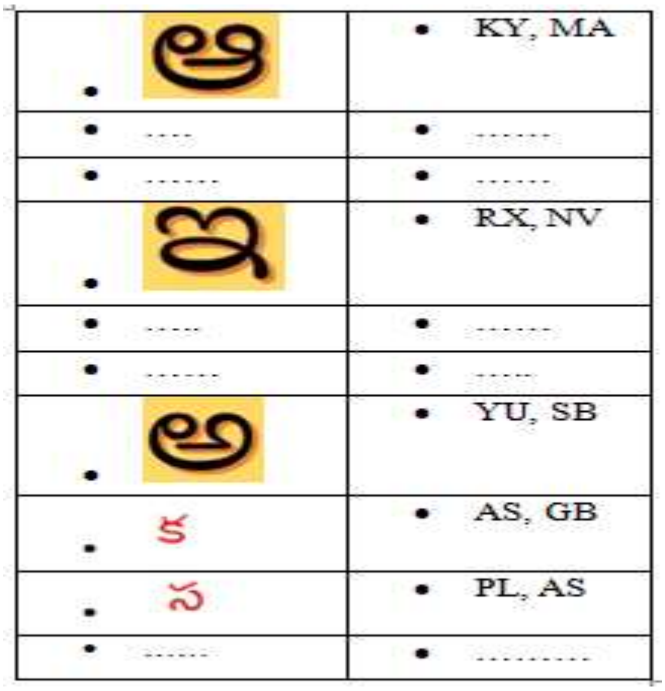
### 3.2 Mapping

Here in mapping each and every Telugu alphabets are mapped to two 2-bit combinations of English alphabets and the result of this phase is called intermediate cipher. Every occurrence of a letter does not have the same mapping there are two 2-bit combination for every Telugu alphabet each occurrence may have one of the two 2-bit combination. The selection of this 2-bit combination of alphabets is made random to make the algorithm effective. It is literally impossible or at least very difficult for the hackers to get the data or text than the traditional algorithms and even though the hacker could crack this mapping he would only get the result of the translation phase i.e. text in Telugu language.

Generating two bit combination of alphabets results in generating 676 two bit characters as shown below. Here the table shows all the two bit combinations of English alphabets. These combinations are used to map Telugu characters which are translated from the plain text. Here in mapping each and every Telugualphabets are mapped to two 2-bit combination of English alphabets and the result of this phase is called intermediate cipher as shown in Figure 3. Every occurrence of a letter does not have the same mapping there are two 2-bit combination for every Telugu alphabet each occurrence may have one of the two 2-bit combination. Mapping is to all the 56 Telugu characters as shown in Figure 4.

| | BA | CA | DA | EA | FA | …. | …. | …. | YA | ZA |
|---|---|---|---|---|---|---|---|---|---|---|
| AB | BB | CB | | | | | | | YB | ZB |
| AC | BC | CC | | | | | | | YC | ZC |
| AE | BD | CD | DD | | | | | | … | … |
| …. | | | | ….. | | | | | …. | …. |
| …. | | | | | ….. | | | | ….. | … |
| ….. | | | | | | | | | YY | …. |
| AZ | BZ | CZ | DZ | …. | …. | ….. | …. | …. | ….. | ZZ |

**Fig.3.** Logic of Mapping Table



| | |
|---|---|
| • ఆ | • KY, MA |
| • ---- | • ------ |
| • ------ | • ------ |
| • ఇ | • RX, NV |
| • ----- | • ------ |
| • ------ | • ----- |
| • అ | • YU, SB |
| • క | • AS, GB |
| • స | • PL, AS |
| • ------ | • ---------- |

**Fig. 4.** Bit Mapping of Telugu characters with English

### 3.3 Encryption and Decryption

This is the final module here we encrypt the intermediate cipher, which is the result of mapping phase using a existing encryption algorithm (AES). Advanced Encryption Standard (AES) is considered more reliable because it uses a 128-bit, 192-bit or 256-bit key. Combining the new algorithm with the existing AES algorithm a highly secure encryption is created. AES algorithm is used since it is more Secure, consumes less memory and more flexible. To hack the data the hacker must decrypt the AES encryption, even though one could be successful in that then 2-bit mapped intermediate cipher must be compromised which is random and changes frequently. If even it was compromised the hacker would only get the translated version (TELUGU TEXT) and must know Telugu to read the actual data.

### 3.4 AES Encryption and Decryption

AES algorithm is an iterated block decipher algorithm with a fixed block size of 128 and a variable key length. The AES algorithm operates on 128 bits of data and generates 128 bits of output. The length of the key used to decrypt this input data can be 128, 192 or 256 bits. AES encryption makes use of four transformations namely substitute bytes, Shift rows, mix columns and add round key. The number of rounds chosen is 10 where each round makes use of all the four aforementioned transformations except the last round. The last round uses only three transformations and omits the mix column transformation. AES decryption too makes use of four transformations used by the encryption algorithm, but in the reverse order. As like encryption the last round uses only three transformations and omits the mix column transformation. Here user A is the sender and user B in the receiver. The plaintext goes through three stages as shown in Fig. 5.

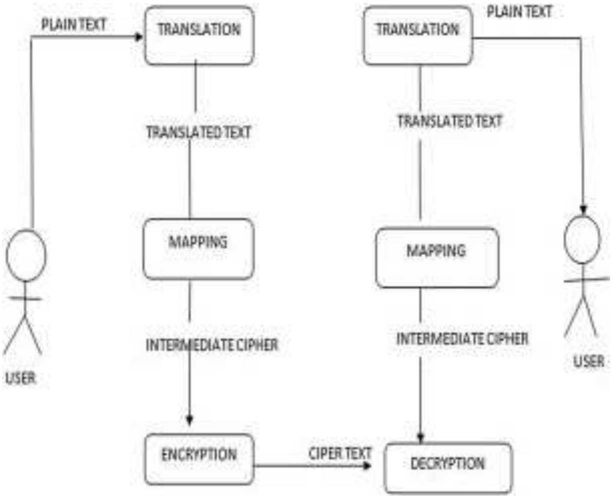


**Fig. 5.** Phases of Telugu Language Encryption

The plain is first translated to Tamil. The translated text is mapped with 2-bit key. The mapped intermediate cipher is encrypted using AES. The process is reversed for decryption. The overall architecture of our proposed model (encryption and decryption) is shown in Fig. 6.

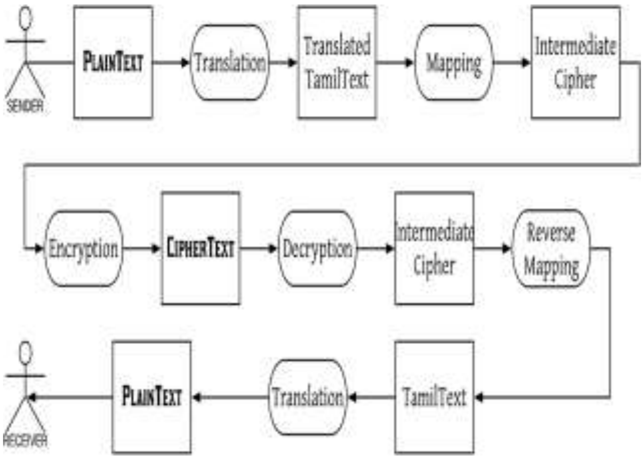


**Fig. 6.** Architecture diagram of Telugu Encryption Model

## IV. PERFORMANCE EVALUATION

The performance of our proposed model has been evaluated using python 3.7.4. Here is the performance evaluation comparison of Telugu Cryptography with DES and Blowfish algorithms in terms of encryption time, memory overhead and avalanche effect.

### 4.1 Encryption Execution Time

Execution time of any cryptographic algorithm involves encryption and decryption timewhich involves changing the plain text to cipher text and vice versa. Since Telugu cryptographyphases like translation and mapping involves translating the plain text from Englishto Telugu language and mapping of Telugu characters to English letters which is a simpleprocess it takes less time to encrypt a file of size 25 kb when compared to DES and Blowfishas shown in Fig. 7.

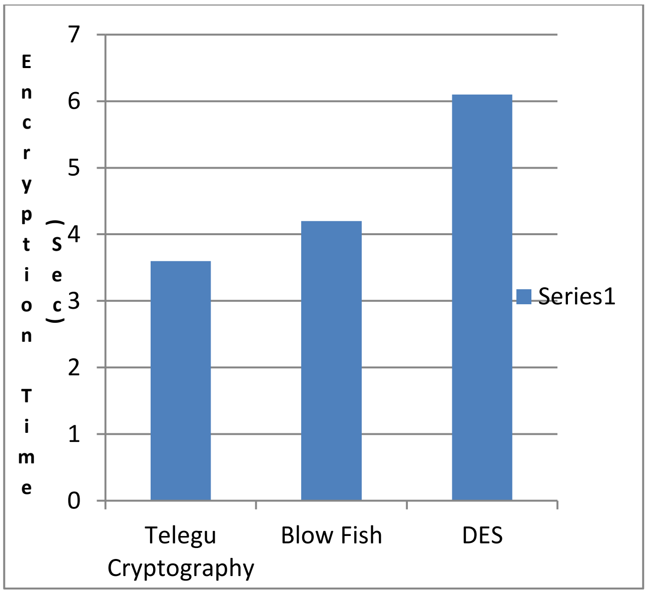


**Fig.7.** Encryption execution time of different algorithms

### 4.2 Memory Usage

Memory allocation is a function assigned to computer programs and services with physical or virtual memory space. Memory allocation is the process of allocating part of the system memory or the complete part of the program to execute program and processes. The memory allocation is achieved through the process called memory management. In storage systems, throughput refers to either the amount of data that can be received and written to the storage medium or read from media and returned to the requesting system, typically measured in mega bytes per second (MBPS). Different encryption algorithms make use of different number of variable as per their requirement for execution thus varies the memory allocation for the same. Here is the detailed memory occupation of different encryption algorithms. Figure 8shows that the memory requirement of our proposed model is comparatively lesser than Blowfish and DES since the translation and mapping phases consumes less space.

### 4.3 Avalanche Effect

Avalanche Effect is one of the desirable properties of any block cipher cryptographic algorithm. It insists the algorithm to change as many number of bits of cipher text as possible if even a single bit of plaintext is changed. A strong avalanche effect is desirable for a good cryptographic algorithm. Avalanche effect is calculated and expressed in percentage as shown in Eq. 1

Avalanche effect (%) = (total number of altered bits in Cipher text/Total number of bits in Cipher text)* 100 Eq. (1)

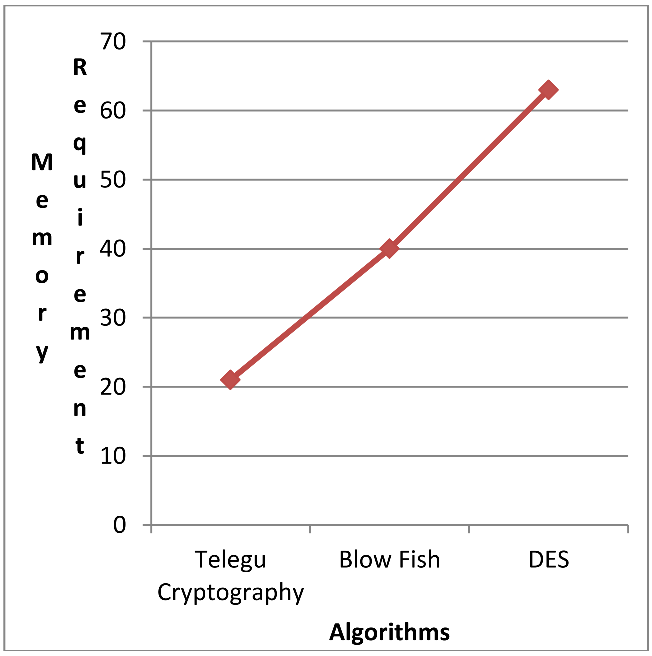


**Fig.8.** Memory requirement of different algorithms

The security of the proposed algorithm is measures with the avalanche effect. Figure 9 shows that our proposed model shows a better avalanche effect for 16 characters plaintext while changing the initial, intermediate and final bit positions since it makes use of AES algorithm when compared to Blowfish and DES.

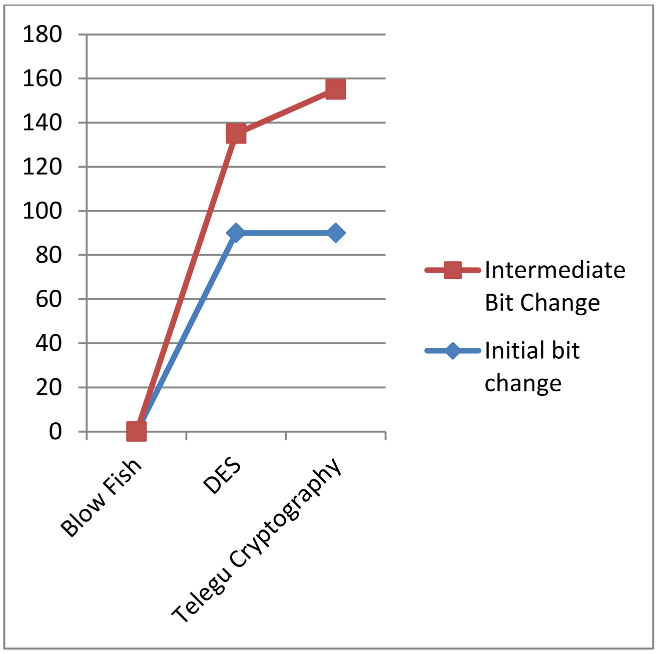


**Fig. 9:** Avalanche effect of different algorithms

## V. CONCLUSION AND FUTURE SCOPE

The development of technology has made people feel insecure about their data.People create millions of data every second and need to be protected from intruders.This article proposed a new efficient hybrid symmetric key cryptographic algorithm called Telugu Cryptography.By encrypting the data using Telugu Cryptography, data transmission over the network can be guaranteed.The evaluation of the proposed algorithm shows that it is superior in terms of memory usage, encryption time, and produces a strong avalanche effect compared to DES and Blowfish algorithms.

## REFERENCES

[1] Geetha R, Padmavathy T, Thilagam T, Lallithasree A. Tamilian Cryptography: An Efficient Hybrid Symmetric Key Encryption Algorithm. Wireless Personal Communications.**17:1-6, Dec. 2019.**

[2] Chaudhary S, Dave M, Sanghi A, Manocha J. An elucidation on steganography and cryptography.InProceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies **4 (pp. 1-6).Mar 2016.**

[3] Odeh A, Elleithy K, Faezipour M. A Reliable and Fast Real-Time Hardware Engine for Text Steganography. IEEE LISAT 2014 Long Island Systems, Applications and Technology. **1-6:2014.**

[4] Reddy PC, Babu AS. A Novel Approach To Analysis District Level Long Scale Seasonal Forecasting Of Monsoon Rainfall In Andhra Pradesh And Telangana. International Journal of Advanced Research in Computer Science.**1;8(9).Nov2017.**

[5] Sucharitha Y, Vijayalata Y, Prasad VK. Analysis of Early Detection of Emerging Patterns from Social Media Networks: A Data Mining Techniques Perspective. InSoft Computing and Signal Processing.**pp. 15-25,2019**. Springer, Singapore.

[6] Al-Haidari F, Gutub A, Al-Kahsah K, Hamodi J. Improving security and capacity for arabic text steganography using ‘Kashida’extensions. In2009 IEEE/ACS International

Conference on Computer Systems and Applications.**10 (pp. 396-399). May2009.**IEEE.

[7] Alkhudaydi M, Gutub A. Securing Data via Cryptography and Arabic Text Steganography. SN Computer Science.**2(1):1-8, Feb. 2021.**

[8] Gutub A, Alaseri K. Hiding shares of counting-based secret sharing via Arabic text steganography for personal usage. Arabian Journal for Science and Engineering.**4:1-26, Jul 2019.**

[9] Baker SI, Al-Hamami AH. Novel algorithm in symmetric encryption (NASE): Based on feistel cipher. In2017 International Conference on New Trends in Computing Sciences (ICTCS) **11 (pp. 191-196) Oct. 2017.**IEEE.

[10] Gupta A, Semwal S, Johari R. METHS: Mapping from English language to Hindi language for secure commercial transactions. In2016 International Conference on Computing, Communication and Automation (ICCCA) **29, pp. 112-116, Apr. 2016.**IEEE.

[11] Vidhya PM, Paul V. A method for text steganography using Malayalam text.Procedia Computer Science.**Jan 1;46:524-31.2015.**

[12] Khairullah M. A novel steganography method using transliteration of Bengali text.Journal of King Saud University-Computer and Information Sciences.**Jul 1;31(3):348-66.2019.**

[13] Taha A, Hammad AS, Selim MM. A high capacity algorithm for information hiding in Arabic text.Journal of King Saud University-Computer and Information Sciences.**2018 Jul 25.**

[14] Hamzah AA, Khattab S, Bayomi H. A linguistic steganography framework using Arabic calligraphy.Journal of King Saud University-Computer and Information Sciences.**2019 May 3.**

[15] Roy S, Venkateswaran P. A text based steganography technique with Indian root. Procedia Technology. **Jan 1;10:167-71.2013.**

[16] Bharati PV, Prasad KJ. Cryptic transmission of Telugu Text. In2016 International Conference on Information Communication and Embedded Systems (ICICES) **Feb 25 (pp. 1-6). 2016.** IEEE.

[17] Changder S, Ghosh D, Debnath NC. Linguistic approach for text steganography through Indian text.In2010 2nd international conference on computer technology and development **Nov 2 (pp. 318-322).2010.**IEEE.

[18] Rao GS, Imanuddin M, Harikumar B. Script Identification of Telugu, English and Hindi Document Image. Int. J. Adv. Eng. Global Technol. **2(2):443-52, 2014.**

[19] Das MS, Reddy CR, Rahul K, Govardhan A. Multilingual Optical Character Recognition System for Printed English and Telugu Base Characters. International Journal of Science and Advanced Technology (ISSN 2221-8386).**Jun;1(4):106-11.2011.**

[20] Shalini M, Indira B. Automatic Character Recognition of Indian Languages–A brief Survey. International Journal of Innovative Science, Engineering & Technology, **Apr;1(2):131-8.. 2014.**

[21] Shaker Reddy PC, Sureshbabu A. An Enhanced Multiple Linear Regression Model for Seasonal Rainfall Prediction. International Journal of Sensors Wireless Communications and Control.**Aug 1;10(4):473-83.2020.**