International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Analysis of Stream Ciphers Based on Theoretic Approach

Shisif Pokhrel[1], Ahmed Abdul Kadhim Basheer[2]

[1]Jawaharlal Nehru Technological University Hyderabad, Andhra Pradesh, Hyderabad, Kukatpally - 500085

[2]Jawaharlal Nehru Technological University Hyderabad, Andhra Pradesh, Hyderabad, Kukatpally - 500085

Abstract: ***This paper is intended to determine the strength of modern security systems by theoretic approach. The design of existing security system is based on complexity of algorithm and secrecy of key. Besides, several parameters exists that may be useful to determine the strength of cryptosystem. Spurious keys gives text like text as decrypted output to a cryptogram and leads to confusion in proceeding towards unique solution during cryptanalysis. A system rich in spurious keys can be considered as more secured system. This paper presents analysis on stream ciphers and provides a construct model with rich spurious keys. The paper also shows the effect of implementation of Natural Language Model in security system***

Keywords: Cryptosystem, Cryptogram, Spurious Keys, Code-Points, Unicity Distance

1. Introduction

Cryptography is a major tool concerned with security of information. The core of cryptography deals with key establishment and secured communication. The secret key is implemented for encryption and decryption. A basic convention exists in cryptography that the algorithm is public, cant be kept hidden to the attacker and the secrecy of the system solely depend on the secrecy of key and complexity of algorithm. However, there are different parameters defined in information theory of secrecy system like Entropy, Redundancy, Unicity Distance, Equivocation, which describe the strength of cryptosystem.

The plain text is converted into cipher text or cryptogram by using encryption algorithm which uses key. When this cipher text is decrypted using same key the original plain text is retrieved. If the cipher text is decrypted using wrong keys the decrypted output is incorrect and usually doesn't look like a text. But, there exists certain keys which give text-like-text as output, called spurious keys. The spurious keys gives output which looks like text but is not correct and provides confusion to attacker. Suppose a phone-number is plain text which is encrypted using certain security system. There exist certain keys which gives phone-number a an decrypted output. These keys provide a dilemma in finding out the original phone-number to adversaries. The modern cryptosystems provide a fair number of spurious keys to short messages .

Basically the messages are based on language and are created using only the characters available in the language model. But, encryption/decryption model does the transformation in bit level which allows each character to transform to any byte value. This leads to larger redundancy in the system and is a reason why the encrypted texts can easily be marked and and the wrong decryption can easily be eliminated during brute-force. Spurious keys are obtained after filtering out the keys which gives output, that does not lie within the text-space specified by language model.

```
plain text:  apple

Actual key:  77201485

Cipher text:  hHri4xg=

Analysis:
{'key: 00253233': 'text_like_message: adhus',
 'key: 01525659': 'text_like_message: lziyv',
 'key: 01654068': 'text_like_message: hrtrh',
 'key: 01893809': 'text_like_message: lmlxf',
 'key: 03764914': 'text_like_message: offec',
 'key: 04757228': 'text_like_message: dexqc',
 'key: 05272803': 'text_like_message: ajdbm',
 'key: 05328929': 'text_like_message: rxwwr',
 'key: 06688108': 'text_like_message: bztad',
 'key: 07491224': 'text_like_message: jodfi',
 'key: 08363491': 'text_like_message: rekin',
 'key: 09007336': 'text_like_message: epbtf',
```

Figure 1: Spurious Keys (ARC4 Algorithm)

Shannon introduced a theory of secrecy system in his paper, “Communication Theory Of Secrecy System” [1], which presents a model of secrecy system in terms of unicty distance and spurious keys. According to Shannon's theory, A system is perfect if all the messages are equiprobable with the cryptogram (ciphertext) intercepted. The cryptogram actually represent a set of messages with a finite number of keys termed as spurious keys. Measurement of the possibility in reaching a unique solution with respect to the amount of cryptogram intercepted is given by Equivocation. There exists threshold to the amount of interception that leads to a unique message, termed as Unicity Distance. Shannon showed that the unicity distance is inversely related to redundancy. Redundancy measures the level of extra bits available in the message with respect to information content.

The cipher text which has length less than Unicity Distance results in number of spurious keys. The number of Spurious Keys is reduced by incorporating language characteristics [2]. Hence, unicity distance is evaluated in terms of Entropy and Redundancy of the Language [3,4]. Higher information rate has lower redundancy, and hence larger unicity distance and ensures encrypted keys, the short cryptogram in a key vault, like Password Safe that cannot be crypt-analyzed [5] within certain limited login attempts. Passphrase with semantic noises has higher information rate, bigger unicity distance, and more spurious keys, which strengthen the login protection with limited attempts. Jeff Hoffstein proposed [6] an Lattice attack on a Spurious key for public key cryptosystem. Rather than trying to find the private key, an attacker might use the lattice (a 2 by 2 matrix composed of four n by n block) and try to find some other short vector in the lattice.

The flaws associated with the existing cryptographic algorithm and the possible construction to achieve ideally stronger secrecy system is presented in the paper. In section 2 trends in cryptosystems is discussed. In section 3, Information Theoretic Approach is introduced and in section 4 and 5 a theoretic analysis is is proposed. Finally, the paper is concluded in section 6.

2. Modern Cryptosystem

Modern cryptosystem implements complex algorithms which is very difficult to execute manually. Besides, the bigger key size like 128 bits for AES, makes it practically impossible to brute-force by implementing available computational resource at present. But lot of records exists where the system thought to be computationally secured were broken badly using different cryptanalysis approach. For example, DES algorithm which was developed in 1970's and adopted by US government was breakable with special hardware. The computational requirement cannot just be a factor to determine the strength of cryptography which modern cryptosystem relies on.

Concerning modern cryptosystem, an encrypted text does not look like a text, they can easily be marked and allows passive attacker to analyze whether the text is encrypted or plain easily. Similarly, when the encrypted text is applied for decryption with wrong keys the output usually doesn't look like a text and so the incorrect key can easily be eliminated. This allows the attacker the understand the cryptogram and come ahead to a solution by eliminating wrong keys.

```
**ARC4 algorithm**
Enter text: This is secret information ...
enter Key: k234ey

cipher:  K������g;�›i    ]?.����$�I�XIR�

Retrieved text:  This is secret information ...

Key: k123ey

Retrieved text:  �i���<�m��;T��\�m���q/J
```

Figure 2: Encryption-Decryption in Modern Cryptosystem

3. Information Theory

A cryptosystem has perfect secrecy if for any message x and any encipherment y, p(x|y)=p(x). This implies that there must be for any message, cipher pair at least one key that connects them. According to Shannon's theory , Suppose a cryptosystem with |K|=|C|=|P|. The cryptosystem has perfect secrecy if and only if each key is used with equal probability of 1/|K|, for every plaintext x and ciphertext y there is a unique key k such that e_k(x)=y.

There are certain issue which has to be addressed theoretically concerning the strength of a cryptosystem which is mentioned in point below:

- The immunity of a system to cryptanalysis when the cryptanalyst has unlimited time and manpower available for the analysis of cryptograms,
- Does a cryptogram have a unique solution (even though it may require an impractical amount of work to find it),
- How much text in a given system must be intercepted before the solution becomes unique,
- Are there systems for which no information could be extracted out whatever is given to the enemy no matter how much text is intercepted.

In the analysis of these problems the concepts of entropy, redundancy, unicity distance and the like developed in "A Mathematical Theory of Communication".

According to Shannon's theory, if the cipher text is equiprobable with any text in plain text-space, the system is perfect. Generally for short messages, there are number of keys associated which gives text like texts on decryption, called spurious keys. The number of spurious keys is a strong factor to determine the strength of the cryptosystem. If the number of spurious keys associated with a message is fairly large then the system tends to perfection with respect to that message text. The number of spurious keys gradually decreases with the size of spurious keys and reaches a negligible value at certain point. Unicity distance is a term which defines threshold after which the possibility of having spurious keys is zero.

4. Text Space Analysis

Plain-text-space is basically defined by the set of characters associated with a language on which texts are written. Generally English alphabets is the concerned text-space that derives English texts. Now, whenever these texts are encrypted, the cryptograms obtained contains foreign elements that are not English alphabets. The reason behind this can be obtained when we observe the construction of the modern cryptographic algorithms. The basic units implemented to transformation of plaintext in cryptography is XOR operation which is applied at bit level. The bit-wise operation leads transformation of a letter to any 1-byte value.

Let us consider, English alphabets (nA=26) are allowed set of characters to plain text. The encrypted/decrypted text can have any 1- byte value(nU=n$\{0,1\}^8$ =256). The probability that the decrypted text is text-like-text is called event. The probability of events of different text size is presented in the Table 3.

Table 3: Probability of Event for English text

S.No.	Text-Size	P.E
1	1	0.1016
2	2	0.0103
3	4	$1.06*10^{-4}$
4	8	$1.13*10^{-8}$
5	16	$1.28*10^{-16}$
6	32	$1.6*10^{-32}$

It can be observed that the probability of events to occur decreases gradually with text size and depends on the plain text space. The probability of event is related to the spurious keys. The observation indicates that the short messages has higher number of spurious keys and so are more resistant to brute-force attack.

5. Analysis Using Natural Language Model

The natural languages needs certain encoding standard to feed to modern cryptosystem for encryption. Utf-8 encoding standard is widely popular as it accepts more then 1 million characters and leave no change on existing ASCII standard characters. The weight of the character codepoints of indian languages are 3-bytes. The heavier weight of the characters of indian scripts gives the possibility of larger universe to encrypted-decrypted texts and thus leading less probability of events to occur. The probability of events to occur based on different text size for Devanagari script is shown in table 4. It is clear that the unicity distance for Devanagari text is approximately 5 which is far less then that of English text.

Table 4: Probability of Event for Devanagari Script implementing UTF-8 encoding

S.No.	Text Size	P.E
1	1	$7.56*10^{-6}$
2	2	$5.73*10^{-11}$
3	4	$3.28*10^{-21}$
4	8	$1.08*10^{-41}$

5.1 Code-point Mapping Technique

Code-point mapping is a method proposed in this paper for fair implementation of natural language models. Basically, code-point mapping techniques works for the languages which has a maximum of 256 character codepoints associated to the script. The codepoint of a character is a number in hexadecimal. This codepoint basically carries two information; information of the language and specificity of the character. For example, the codepoint value of a character 'ka' →' □' is 0x0915. Here '0x09' part of the number remains through out the script and '15' specifies the character. So, the part ('15') of the number can be mapped to a 1-byte value ('\x15'). The mapped value is encrypted and the extracted part ('0x09') can be appended back to the encrypted data to obtain the cipher. The block diagram of encryption is shown in figure 3 and implementation is presented in figure 4 and 5.

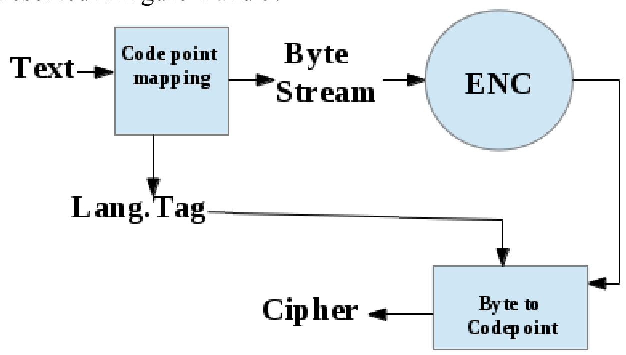

Figure 3: Block Diagram of Encryption scheme using codepoint Mapping

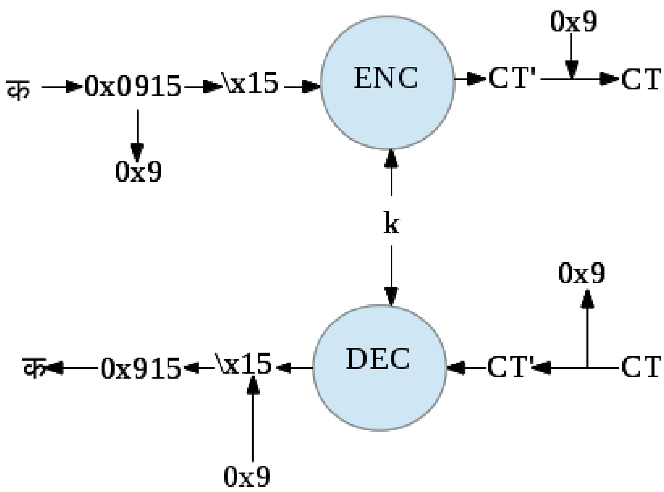

Figure 4: Implementation of code-point Mapping on Encryption and Decryption

The probability of event for Devanagari text after the implementation of codepoint mapping, is fairly larger than for English. Table 5 provides the proof of concept for probability of events for Devanagari Script. The unicity distance where the probability of event becomes negligible (less than 2^{-80}) is 82 which is fairly greater compared to English

Table 5: Probability of spurious keys implementing Devanagari Language

S.No.	Text Size	P.E	ARC4	DES
1	4	0.0606	0.0603	x
2	8	$3.7*10^{-3}$	$3.6*10^{-3}$	$3.6*10^{-3}$
3	16	$1.3*10^{-5}$	$1.3*10^{-5}$	$1.6*10^{-5}$
4	32	$1.8*10^{-10}$	_	_

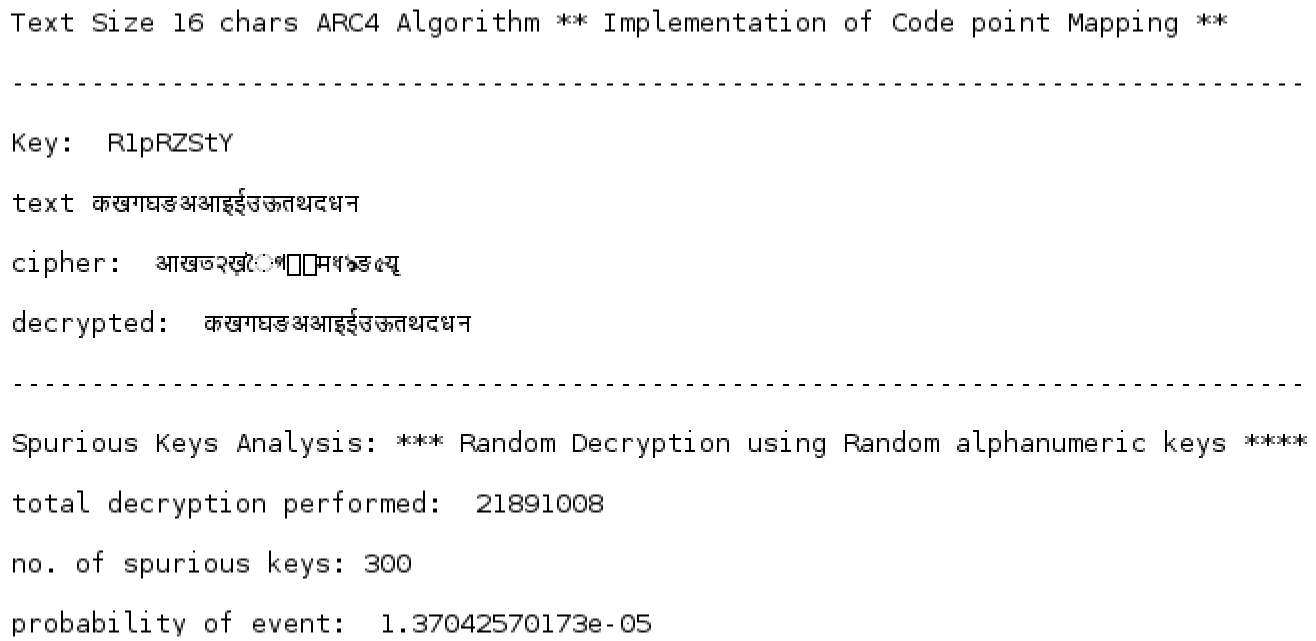

Figure 5: Spurious key Analysis on natural Language Model implementing Code-point Mapping

6. Construction for Stream Cipher

Stream ciphers is the practical application of one-time-pad. It consist of a Pseudo Random Generator (PRG) which takes the supplied key as a seed to the pseudo random generator. The PRG generates a long bit sequence which is equal to the text to be encrypted. The encryption is simply the one-time-pad of the text with generated bit stream. The one-time-pad holds a property with respect to the text space which can be illustrated with simplicity by considering a text-space with only two element '\x00' and '\x01' ie. U= $\{\backslash x00, \backslash x01\}^n$ where n is the size of text. The encrypted and decrypted texts can also be bounded to the same space-limit by applying MOD-2 operation to the code-points after XOR. The property holds true for text space of with set of elements of size 2^x, x= 1 to 8, where MOD-2^x is applied to bound the encrypted-decrypted texts.

6.1 Mathematical Modeling

Let us define a simple model simple encryption and decryption model for stream cipher by equation (1) and (2) where m,c,d,k are respectively message, cipher, decrypted text, key and E(), D() be the encryption and decryption algorithm function which takes two arguments as given.

$$c = E(m,k) \qquad \text{---------------(1)}$$

$$d = D(c,k) \qquad \text{------------------(2)}$$

For stream cipher E(m,k) and D(c,k) may be defined as below

$$E(m,k) = m_b \text{ OTP } K_b \qquad \text{--------(3)}$$

where K_b =G(k) and G() is a Pseudo Random Generator which takes key k and generates K_b

$$D(c,k) = c_b \text{ OTP } K_b \qquad \text{-------------- (4)}$$

This implies,

$$D(E(m,k),k) = m \qquad \text{-------------------(5)}$$

If a plain text of size n-byte be defined by m, where m∈ U=$\{\backslash x00,\backslash x01 \ldots\}^n$ and size of the set $\{\backslash x00,\backslash x01 \ldots\}$ be given by 2^p, p ∈{1,2,...8} , then

$$c=E'(m,k) = [\ (x \text{ MOD } 2^p) : x = \text{byte-value for each byte in } E(m,k)\ ] \qquad \text{--------(6)}$$

$$D'(c,k) = [\ (x \text{ MOD } 2^p) : x = \text{byte-value for each byte in } D(c,k)\ ] \qquad \text{-----------(7)}$$

case-1:
For p =8,

$$E'(m,k) = E(m,k) \qquad \text{----(8)}$$

D'(m,k)=D(m,k) ------(9) and hence, the model replicates the original stream cipher system with p=8

case-2:
For p=0,

$$E'(m,k) = D'(m,k) \qquad \text{-------(10)}$$

and hence, this does not follow general rule of crpytography.

The probability of collision increases with smaller value of p along with the condition that with larger text size (n) the probability of collision is depreciated. Thus, with possible tradeoffs this model can be implemented to attain large spurious keys.

6.2 Index Mapping

Index mapping is an approach of implementation of stream cipher with boundary to text space. The implementation of boundary to text in stream cipher allowed anly ist 2^P bytes ,p ∈ {1,2...8}, to appear in text space. For example for for p = 2, only ist 2^2 bytes i.e '\x00', '\x01' , '\x02', '\x03' are allowed elements to text-spaces. These number of bytes can be mapped to the indices of equal set of desirable characters and implemented encryption. An example for p=3 is presented in Figure 6.

```
** text space boundary in Stream Cipher (ARC4) **

Characters space:  ['1', '2', '3', '4', '5', '6', '7', '8']

**Index Mapping**
1 --> 0 --> \x00
2 --> 1 --> \x01
3 --> 2 --> \x02
4 --> 3 --> \x03
5 --> 4 --> \x04
6 --> 5 --> \x05
7 --> 6 --> \x06
8 --> 7 --> \x07

Plain Text:  132423525454645685734858238528283842348428314823843438276486438563866283462846814138432853853

Key:  cnlhxckq

Cipher Text:  771375752623827444322451553175282171138761585711387216182438171658156673174166665774547216442

Retrieved Text:  132423525454645685734858238528283842348428314823843438276486438563866283462846814138432853853
```

Figure 6: Index Mapping with Text-Space of 8 characters

The numeric string or numeric text is composed of 10 characters belonging to set string.digits i.e. ['0','1','2','3','4','5','6','7','8','9']. An encryption scheme with 8 numeric characters set with boundary to text space is mentioned in the above. There are total ${}^{10}C_8$ combinations of 8 numeric characters set. The text space string.digits can be realized as ${}^{10}C_8$ =45 unique text-spaces with 8-numeric characters. The implementation of 45 encryptions with boundary to those 45 text-spaces would result an equivalent encryption to numeric strings with encrypted-decrypted texts bounded to string.digits. Figure 7 shows the implementation of numeric string encryption with boundary.

```
***NumericStrings Encryption***
*** Plain Text is the list of reported Phone Numbers in March 2014***

Characters space: ['0', '1', '2', '3', '4', '5', '6', '7', '8', '9']

Plain Text:    2317654558  4903658313  8932321335  2345306027  15[illegible]
010  1636464037  2763381796  8333123812  4824820911  8339248602  1542904384  7865457672  8739229048  3583277802  66468
25637  4174333068

Key:  wmphopoy

Cipher Text:    6576135597  1774935993  3411214361  8254735186  18[illegible]  1443532805  400312
7395  6759605756  6621607725  1230301488  1548538322  8956543607  3918214219  7189874357  8774407379  2611852601  3912
188849  1834061161

Retrieved Text:    2317654558  4903658313  8932321335  2345306027  1502664559  6856892790  1616550196  7685391091  482
3144010  1636464037  2763381796  8333123812  4824820911  8339248602  1542904384  7865457672  8739229048  3583277802  6
646825637  4174333068
```

Figure 7: Numeric String Encryption Model

7. Conclusions and Future Scope

The analysis of spurious keys provides a vision of strengthening cryptography by leading cryptosystems beyond brute-force bound. Some of the points were observed with statistics drove by an approach of random and brute-forced decryptions as a proof of concept in this research work. The analysis was done in modern cryptographic algorithms which include both block ciphers (DES, DES3, Blowfish) and stream cipher(ARC4).

- The probability of spurious keys to occur at random decryption for 8-byte texts with text space, 26 English alphabets is 10^{-8} whereas for 8 characters text from Devanagari Script text is $3*10^{-3}$ when code-point mapping is implemented for encryption.
- The probability of event decreases gradually with text size and is negligible at certain point, unicity distance. Considering 2^{-80} as a negligible probability, the unicity distance for 26 alphabet text is approximately 27 characters while that for Devanagari text round offs to 81 characters.
- The property of One Time Pad indicated that there is a possibility of bounding encrypted-decrypted texts by limiting number of allowed characters in stream ciphers. An encryption model is designed based on the property of OTP for numeric string for which the unicity distance tends to infinity.

Detailed analysis with generation of spurious keys has to be carried out to quantify the quality of secrecy system and evaluate the theoretic approach proposed in this paper.

References

[1] C. E. Shannon (1948) " A Mathematical Theory of Communication," In The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656
[2] Shyue-Ching Lu (1979) "The Existence of Good Cryptosystems for Key Rates Greater than the Message Redundancy," In IEEE Transactions on Information Theory, Vol. IT-25, No.4, pp 475-477
[3] Martin E Hellman (1977) "An Extension of the Shannon Theory Approach to Cryptography," In IEEE Transactions on Information Theory, Vol. IT 23, NO. 3, pp. 289-294
[4] Rolf J. Blom (1984) "An Upper Bound on the Key Equivocation for Pure Ciphers," In IEEE Transactions on Information Theory, VOL. IT 30, NO. 1, pp 82-84
[5] Kok-Wah Lee, Hong-Tat Ewe (2007) "Passphrase with Semantic Noises and a Proof on Its Higher Information Rate," in International Conference on Computational Intelligence and Security Workshops, pp-652-655
[6] Jeff Hoffstein, Daniel Lieman, Jill Pipher, Joseph H. Silverman "NTRU: A Public Key Cryptosystem," NTRU Cryptosystems, Inc., pp. 11-14
[7] Bhadri Raju MSVS, Vishnu Vardhan B, Naidu G A , Pratap Reddy L, Vinaya Babu A: Effect of Language Complexity on Deciphering Substitution Ciphers - A Case Study on Telugu.
[8] Pratap Reddy L: A New Scheme for Information Interchange in Telgu through Computer Networks : Doctoral Thesis. JNTU,Hyderabad, India, 2001.
[9] Adam Stone, Internationalizing the Internet. J. Internet Computing. 3, 2003, pp. 11-12. Adam Stone, Internationalizing the Internet. J. Internet Computing. 3, 2003, pp. 11-12.
[10] U.Maurer (1999) "Information-Theoretic cryptography," In Advances in Cryptology-CRYPTO'99 In Lecture Notes in Computer Science, Springer-Verlag, vol. 1666, pp. 47-64
[11] Zhaozhi Zhang (2005) "A Simplified Method for Computing the Key Equivocation for Additive-Like

Instantaneous Block Encipherer," In J. Electronic Notes in Discrete Mathematics, Vol. 21, pp. 389–391

[12] Bhadri Raju MSVS, Vishnu Vardhan B, Naidu G A, Pratap Reddy L and Vinaya Babu A (January, 2012) "Effect of Language Complexity on Substitution Ciphers- A case study on Telugu," Internation Journal of Security and Its Applications, Vol. 4, No. 1, pp 11-20

Author Profile

Shisif Pokhrel has received the B.E. Degree in Electronics and Communication Engineering (ECE) from Advanced College of Engineering and Management (ACEM) affiliated to Tribhuwan University, Nepal in the year 2011 and is pursuing M.Tech. degree in Embedded Systems under Department of ECE from Jawaharlal Nehru Technological University Hyderabad (JNTUH). He is working in projects related to Network and Information Security.

Ahmed Abdul Kadhim Basheer completed his BTech at Technical College of Najaf, Iraq. He is persuing his MTech towards Digital Systems and Computer Engineering at JNTUH, India. His research areas are towards Wireless Communication, Network Security, VLSI and Robotics.