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.

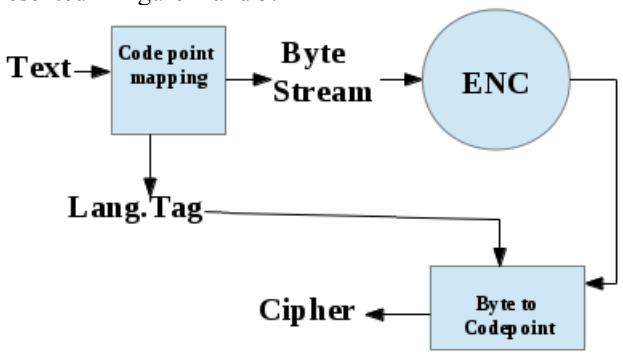

Figure 3: Block Diagram of Encryption scheme using codepoint Mapping

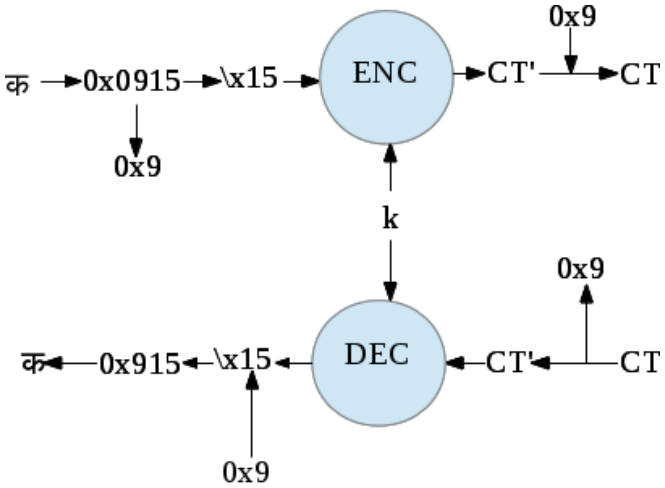

Figure 4: Implementation of code-point Mapping on Encryption and Decryption

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	$1.8*10^{-10}$	_	_

```
Text Size 16 chars ARC4 Algorithm ** Implementation of Code point Mapping **

--------------------------------------------------------------------------------

Key:  RlpRZStY

text कखगघङअआइईउऊऋतथदधन

cipher:  आखङ२ख़ँ◌ग□□मधঌङৎयू

decrypted:  कखगघङअआइईउऊऋतथदधन

--------------------------------------------------------------------------------

Spurious Keys Analysis: *** Random Decryption using Random alphanumeric keys ****

total decryption performed:  21891008

no. of spurious keys: 300

probability of event:  1.37042570173e-05
```

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= $\{\x00,\x01\}^n$ 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.

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

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

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

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

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

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

This implies,

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

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

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

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

case-1:
For p =8,

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

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

case-2:
For p=0,

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

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 $\in\{1,2...8\}$, 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.

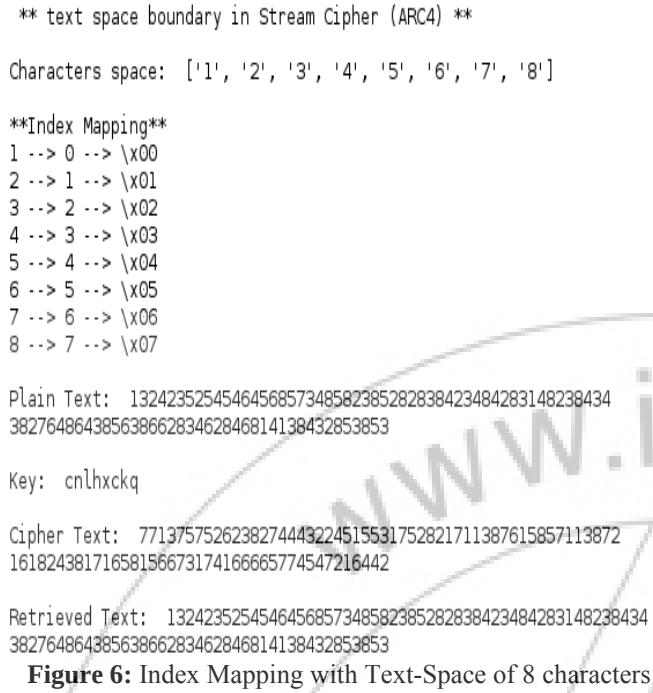

Figure 6: Index Mapping with Text-Space of 8 characters

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.

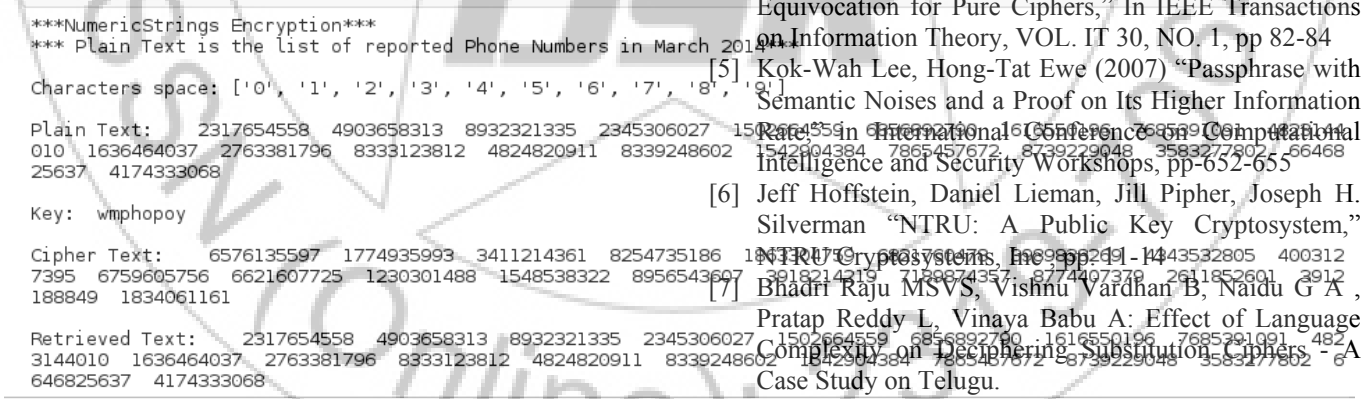

Figure 7: Numeric String Encryption Model

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. 1 - 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.