Selective Encryption using Natural Language Processing for Text data in Mobile Ad hoc Network

Ajay Kushwaha[1], Hari Ram Sharma[2] ,Asha Ambhaikar[3]
{ajay.kushwaha@rungta.ac.in[1], hrsharma44@gmail.com [2] ,dr.asha.ambhaikar@rungta.ac.in[3] }

Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India[123]

Abstract. These days security is highly recommended for data which is transferred over network. Although various approaches towards security are proposed day-by-day, the security loop holes are also propagating and thus constant innovation is essential towards protection of data. The paper aims to introduce a selective encryption algorithm for text data encryption termed as Selective Significant Data Encryption (SSDE) adding a noteworthy uncertainty to data while encryption. The SSDE provides sufficient uncertainty to the data encryption process as it selects only significant data out of the whole message using Natural Language Processing (NLP). This in turn reduces the encryption time and enhances the performance. Symmetric key algorithms are typically efficient and fast cryptosystem over other methods, So BLOWFISH method is used for encryption. The SSDE method is found superior to existing ones, that is, toss-a-coin and full encryption method when performance is evaluated based on the extensive set of experiments.

Keywords: Cryptography, Mobile Ad hoc Network, Selective Encryption, Selective Significant Data Encryption, Natural Language Processing, Stop Words.

1 Introduction

The world is moving towards wireless network nowadays and thus ad hoc networks are also acquiring importance. An Ad hoc network is defined as a wireless network in which, all the nodes are able to communicate with each other directly without the need of a central access point. The performance of Ad hoc network is good when less number of nodes are involved, but when the number of nodes increases, the performance gets affected and becomes difficult to manage. The mobile features make the nodes in the Ad hoc network moving. As mobile ad hoc networks is widely used nowadays so the security requirements for the network is also increasing which can be provided by means of cryptography.

There may be two ways of keeping information secret: one is hiding the existence of the information and second is making the information unintelligible. Cryptography could be defined as the art and science of making the information secure from unintended audiences by encrypting it and thus making it unintelligible. Conversely, Cryptanalysis is the art and science of decrypting the encrypted data. The plain text is converted to cipher text while performing encryption and the cipher text is converted back to plain text in decryption. This cipher text is unintelligible to others while being transmitted in the network.

The Encryption and Decryption could be performed by the use of keys. There are two types of key-based encryption, symmetric and Asymmetric algorithms. In case of symmetric algorithms, the key is same for both encryption and decryption, while asymmetric algorithms possess different keys. Symmetric algorithms may have stream ciphers and block ciphers. Stream ciphers encrypts single bit of plain text at a time, whereas block ciphers encrypts a number of bits of plain text as a single unit. The key here is called secret key used in both sender's as well as receiver's end. One of the examples of symmetric algorithm is DES.

In Asymmetric algorithm, the public key is available at both ends while private key is available at only one side. When data is encrypted by public key, it can be decrypted by only private key and vice versa. The algorithm also called as public key cryptography, provides the fit of authenticating the source as a means of digital signature. An example for asymmetric algorithms is RSA.

Natural Language Processing is related to the field of computer science, artificial intelligence and computational linguistics concerned with computers and human languages. Natural Language Processing is related to the area of interaction between humans and computers. The biggest challenges in Natural language processing are natural language understanding, word processing, Information management and enabling computers to obtain meaning from humans.

Selective encryption algorithms are popular in current scenario is due to the fact that they may reduce the overhead spent on data encryption/decryption, and thus improve the efficiency of the network. This whole task is performed with the help of NLP. The approach removes the stop words from the messages and encrypts the significant data only, prior to sending over the network. The stop words are those words which are filtered out prior to or after from natural language text. They are common words which would to be of little value to the messages.

The paper is organized in following way: Related Work is provided in section 2. Section 3 gives the concept of Selective Encryption, Proposed method SSDE is introduced in section 4, which is followed by the Result analysis in section 5. The paper is concluded in section 6.

2 Related Work

Yonglin et al [1] present a probabilistic selective encryption algorithm which utilizes the advantages of the probabilistic methodology that aims to acquire additional uncertainty in text.

Matin et al. [2] examine the performance of the new cipher in MANET and wireless LAN networks and make a performance comparison with that of AES. In the paper, they focused on the security that is provided at the application level. As the key size of the algorithm is larger, the time required to break an encryption scheme becomes so excessive that undesirable attacks are meaningless.

Shivendra and Aniruddha [3] propose and implement a combined approach for identification of a given unknown sample of cipher text. In the first part of system, cipher text samples are generated randomly using different cipher algorithms. In the second part; the system analyses sample through a) Block Length/stream Detection b) Entropy/Reoccurrence Analysis c) Dictionary and Decision tree based approach.

Zhou and Tang [4] proposed an implementation of a complete and practical RSA encrypt/decrypt solution based on the study of RSA public key algorithm. In addition, the encrypt procedure and code implementation is provided in details.

Umaparvathi and Varughese [5] presents a comparison of the most commonly used symmetric encryption algorithms AES (Rijndael), DES, 3DES and Blowfish in terms of power consumption. A comparison has been conducted for those encryption algorithms at different data types like text, image, audio and video. Experimental results are given to demonstrate the effectiveness of each algorithm.

Chang et al. [6] presents a powerful and versatile security suite for the AODV (Ad-hoc On-demand Distance Vector) routing protocol. It offers coverage on common security aspects such as encryption and authentication, and it can be easily modified to work with any distance-vector-based routing for MANET (Mobile Ad hoc Networks). The suite utilizes powerful authentication and user-adjustable encryptions based on digital certificate chaining and popular ciphers such as DES, AES, and RSA.

Nawneet et al. [7] is concluded by a comprehensive summary which discussed the vulnerabilities, challenges and security attacks on ad-hoc routing protocols which leads to difficulties in designing and development of a secure routing protocol and is challenging task for researcher in an open and distributed communication environments.

The security mechanisms presented in this paper by Michiardi and Molva [8] are a practical response to specific problems that arise at a particular layer of the network stack. However, the proposed solutions only cover a subset of all possible threats and are difficult to integrate with each other.

Suresh et al. [9] study the major attack types that MANET faces and the security goals to be achieved. This paper gives out a brief survey of major security protocols with their relative comparison.

3 Concept of Selective Encryption

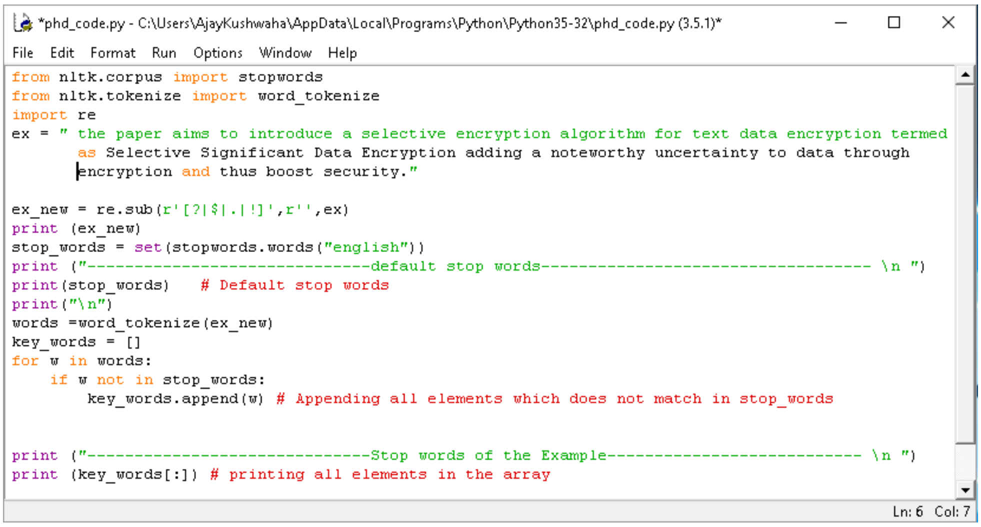
Selective encryption algorithms are popular in current scenario is due to the fact that they may reduce the overhead spent on data encryption/decryption, and thus improve the efficiency of the network. In this section, we have proposed an algorithm for selective encryption and some commonly used techniques of the selective encryption.

The purpose of selective encryption algorithms is to encrypt only certain portions of the messages and provide trustworthy safety so as to secure the transmitted message confidentiality. Selective encryption is proficient to improve the scalability of data transmission and also reduces the processing time. NLP is used for selective encryption of messages. Natural Language Tool Kit (NLTK) 3.0 with python 3.5 version is used for analysing messages under this proposed method [10].

3.1 Steps of selective encryption while processing messages:

1. Removing special characters from the messages like *$&? Etc.
2. Process of tokenization in which it extracts all the words present in the messages
3. Dropping stop words (common words) and collecting significant data (key words) from the messages
4. All key words are encrypted and rest common words are send as it is on to the network.

In order to understand the steps of selective encryption, code along with an example is given in Figure 1 (a) & (b).

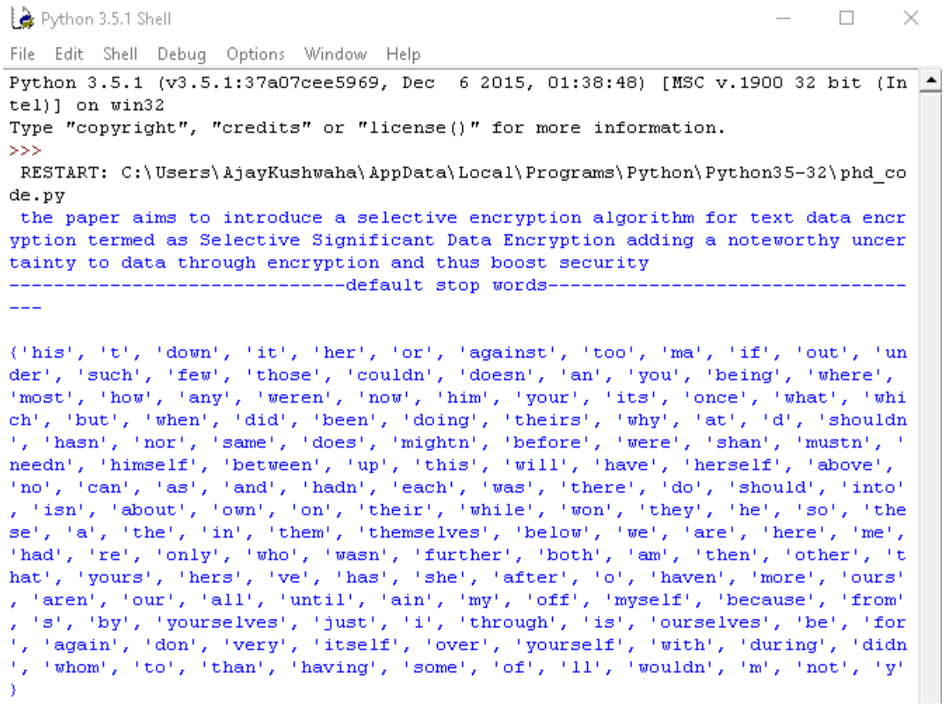

phd_code.py - C:\Users\AjayKushwaha\AppData\Local\Programs\Python\Python35-32\phd_code.py (3.5.1)
File Edit Format Run Options Window Help
Ln: 6 Col: 7

```python
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import re
ex = " the paper aims to introduce a selective encryption algorithm for text data encryption termed
        as Selective Significant Data Encryption adding a noteworthy uncertainty to data through
        encryption and thus boost security."

ex_new = re.sub(r'[?|$|.|!]',r'',ex)
print (ex_new)
stop_words = set(stopwords.words("english"))
print ("------------------------------default stop words----------------------------------- \n ")
print(stop_words)    # Default stop words
print("\n")
words =word_tokenize(ex_new)
key_words = []
for w in words:
    if w not in stop_words:
        key_words.append(w) # Appending all elements which does not match in stop_words


print ("------------------------------Stop words of the Example--------------------------- \n ")
print (key_words[:]) # printing all elements in the array
```

Fig. 1(a) Python code along with NLTK for word processing



```
Python 3.5.1 (v3.5.1:37a07cee5969, Dec  6 2015, 01:38:48) [MSC v.1900 32 bit (In
tel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>> 
 RESTART: C:\Users\AjayKushwaha\AppData\Local\Programs\Python\Python35-32\phd_co
de.py
 the paper aims to introduce a selective encryption algorithm for text data encr
yption termed as Selective Significant Data Encryption adding a noteworthy uncer
tainty to data through encryption and thus boost security
------------------------------default stop words--------------------------------
---

{'his', 't', 'down', 'it', 'her', 'or', 'against', 'too', 'ma', 'if', 'out', 'un
der', 'such', 'few', 'those', 'couldn', 'doesn', 'an', 'you', 'being', 'where',
'most', 'how', 'any', 'weren', 'now', 'him', 'your', 'its', 'once', 'what', 'whi
ch', 'but', 'when', 'did', 'been', 'doing', 'theirs', 'why', 'at', 'd', 'shouldn
', 'hasn', 'nor', 'same', 'does', 'mightn', 'before', 'were', 'shan', 'mustn', '
needn', 'himself', 'between', 'up', 'this', 'will', 'have', 'herself', 'above',
'no', 'can', 'as', 'and', 'hadn', 'each', 'was', 'there', 'do', 'should', 'into'
, 'isn', 'about', 'own', 'on', 'their', 'while', 'won', 'they', 'he', 'so', 'the
se', 'a', 'the', 'in', 'them', 'themselves', 'below', 'we', 'are', 'here', 'me',
'had', 're', 'only', 'who', 'wasn', 'further', 'both', 'am', 'then', 'other', 't
hat', 'yours', 'hers', 've', 'has', 'she', 'after', 'o', 'haven', 'more', 'ours'
, 'aren', 'our', 'all', 'until', 'ain', 'my', 'off', 'myself', 'because', 'from'
, 's', 'by', 'yourselves', 'just', 'i', 'through', 'is', 'ourselves', 'be', 'for
', 'again', 'don', 'very', 'itself', 'over', 'yourself', 'with', 'during', 'didn
', 'whom', 'to', 'than', 'having', 'some', 'of', 'll', 'wouldn', 'm', 'not', 'y'
}
```

Fig. 1(b) Example of word processing

In case of selective encryption algorithms, there is involvement of uncertainty in the message encryption process while determining the uncertain pattern of encrypted messages. Thus, uncertainty may enhance the security of data transmission, since all messages are assumed to have equal importance. Thus, uncertainty becomes one of the principle factors while designing a selective-encryption based cryptosystem. Usually, the more is the uncertainty involved, the more is the cryptosystem effective.

At the present time, selective encryption algorithms are mainly applied in the energy-aware environments or large- scale data transmission, such as, Wireless Sensor Networks (WSNs), Mobile Ad Hoc Networks (MANETs), multimedia communications, etc. In a WSN, each device uses battery as its power supply and therefore has inhibited computational ability, so it is difficult for a sensor to spend too much computational cost on data encryption and decryption. Under these circumstances, the design of a selective encryption algorithm with less processing time but with comparatively high security level is enormously significant. Multimedia communication often requires real-time data transmission, so large amount of audio and video data need to be transmitted securely. If all multimedia data are encrypted, this will create large amount of overhead, so multimedia data is difficult to transmit timely and the quality of communication cannot be guaranteed.

3.2 Full Data Encryption

In Full Data Encryption whole data that is to be sent over the network is encrypted before transmitted to the receiver side.

3.3 Toss-A-Coin Method

This method is a form of Selective Encryption, in which the whole message which is to be transmitted is divided into two groups- even and odd and from the starting of the message, each odd word belongs to odd group and each even word belongs to the even group. The uncertainty involved here is which group will be encrypted i.e. even or odd is not known. As only one group is encrypted, it makes the encryption selective. Now which group will be encrypted is decided by tossing a coin. Here only 50% data is encrypted, thus not much data is reduced and also involvement of uncertainly is less.

4 Selective Significant Data Encryption (SSDE)

The approach selects the significant data there in the message and encrypts them prior to sending over the network. Significant data implies the keyword that holds the meaning of entire message. Excluding significant ones, rest commonly used words like articles, pronouns, conjunctions, prepositions, and interjections are sent without encoding. The flowchart of the proposed method (i.e., SSDE) is given in Figure 2 which shows the execution of SSDE algorithm.

Proposed Algorithm

Step1. Input Messages
Step2. Remove special characters from the messages like *$&? Etc
Step3. Process of tokenization in which it extracts all the words present in the messages.
Step4. Dropping stop words (common words) and collecting significant data (key Words) from the messages using NLTK package in python.
Step5. Encrypt significant data (key words) using Blowfish algorithm and send the message to network.
Step6. Send the stop words to network without encryption.

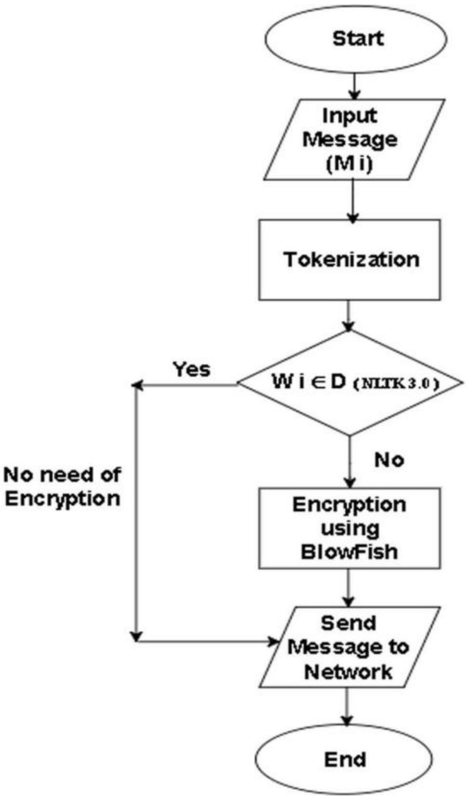


Fig. 2 Flow chart of Selective significant data encryption

5 Result Analysis

In order to observe the characteristics of SSDE, we carried out an extensive set of experiments within a wireless environment. The experimental setup is done using Red Hat 6.0 32-bit Operating System and NS 2.34, processor: Intel(R) core(TM) i3 CPU M 480 @ 2.67 GHz, 2.66 GHz, 4 GB installed RAM. In this work, the proposed method SSDE is compared with commonly used techniques, that is, Full Encryption and Toss-A-Coin method. Each experiment is run for 50ns of simulation time. During the simulation experiment, the compared systems are all run under the identical scenario. The performance metrics for evaluating the SSDE are Encryption Time, Decryption Time and Battery consumption.

The symmetric key encryption algorithms AES, DES and BLOWFISH are implemented using Red Hat 6.0 32-bit Operating System and NS 2.34 with packet size for text data to be 512 bytes. The experimental result shows that Blowfish performs better for throughput. So, we will be using Blowfish algorithm for our proposed method.

As stated earlier, two approaches are used as the comparable models with our proposed system. The first approach encrypts all messages without leaving any text unencrypted and thus termed as Full Encryption. In the second approach half of the data is encrypted and is termed as Toss-a-Coin method.

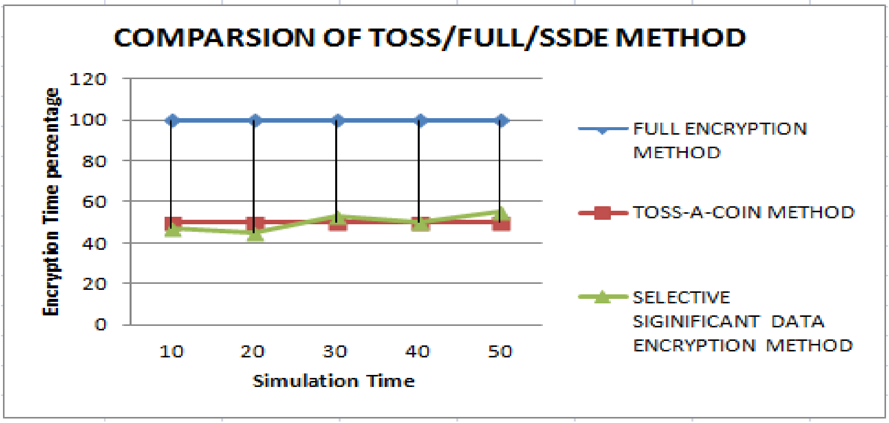


Fig. 3 Encryption Time Percentage vs. Simulation Time

Figure 3 and 4 represent the comparison of encryption, decryption and simulation time based on three approaches. Figure 3 show that both toss-a-coin and SSDE have an obvious lower encryption time than full decryption. This advantage is because of selective encryption which reduced the overhead. In Figure 4 the decryption time in full encryption is more as compared to both toss-a-coin and SSDE and thus selective encryption is superior to full encryption for utilization of resources in the network.

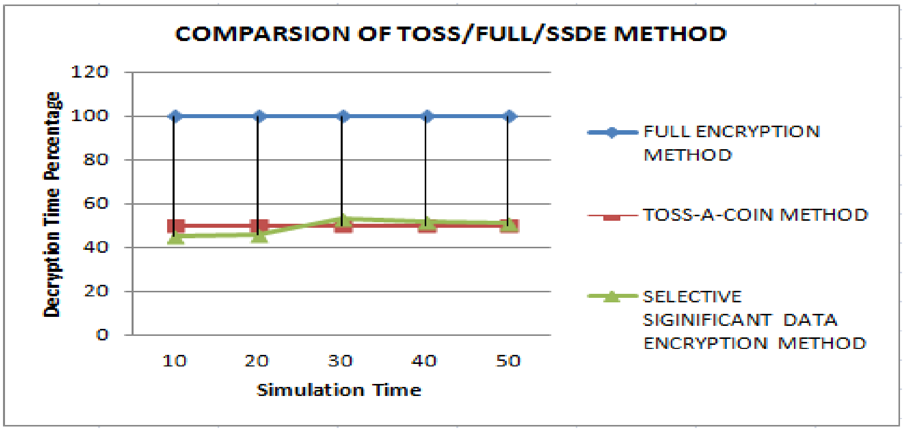


Fig. 4 Decryption Time Percentage vs Simulation Time

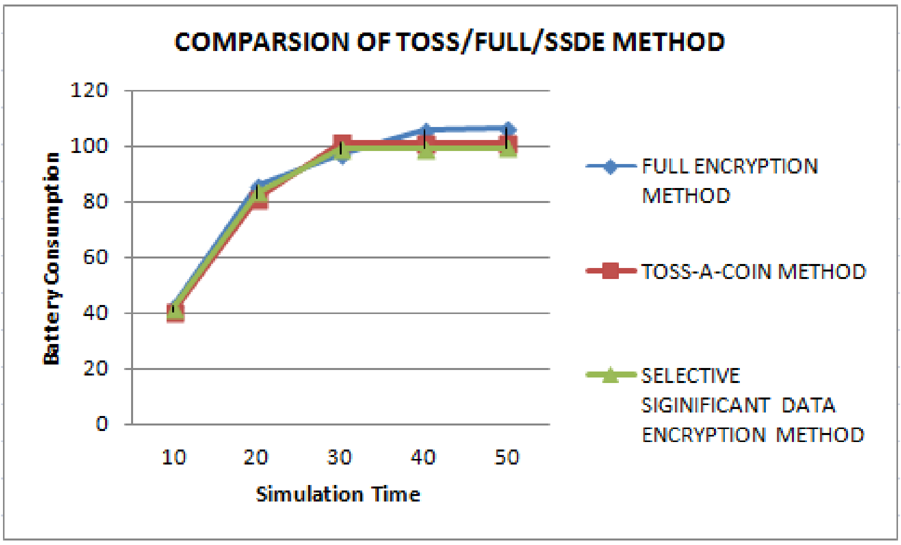


Fig. 5 Battery Consumption vs. Simulation Time.

Figure 5 compare the battery consumption of toss-a-coin, full and SSDE respectively. Figure 5 displays that SSDE has lower battery consumption than full method and more than toss-a-coin. As it is difficult to identify what parts of messages are encrypted in SSDE, thus gives an added advantage. It is evident that SSDE is more efficient and time saving when compared with full and toss-a-coin method in all aspect like encryption ,security etc.

6 Conclusion

This paper introduces a better solution for data encryption in wireless networks. The approach is based on Selective encryption, which is one of the most promising solution nowadays to reduce cost of data protection as well as providing sufficient uncertainty for reliability and improved data security. The performance of the method is evaluated based on the extensive set of experiments. The results demonstrate the effectiveness of SSDE over other methods in wireless networks. Thus the provided solution gives a feasible solution for secure wireless communication in Mobile Ad hoc network. This method can be used in social chatting apps, military security, corporate world communication, and government activities involving text data encryption. This method can be used for text data only. In future, this method can be extended for other file formats (i.e. Audio, video etc).

References

[1] Boukerche, A., Mokdad, L., Ren, Y.: Performance Analysis of a Selective Encryption Algorithm for Wireless Ad hoc Networks. In: Wireless Communications and Networking Conference, pp. 1038- 1043. IEEE, (2011)

[2] Matin, M.A., Hossain, M. M., Islam, M.F., Islam, M.N., Hossain, M.M.: Performance Evaluation of Symmetric Encryption Algorithm in MANET and WLAN. In: International Conference for Technical Postgraduates (TECHPOS), pp. 1 – 4. IEEE , Kuala Lumpur , (2009)

[3] Mishra, S., Bhattacharjya A.: Pattern Analysis of Cipher Text: A Combined Approach. In: International Conference on Recent Trends in Information Technology (ICRTIT), pp. 393 – 398. IEEE,(2013)

[4] Zhou, X., Tang, X.: Research and Implementation of RSA Algorithm for Encryption and Decryption. In: International Conference on Strategic Technology (IFOST), pp. 1118 – 1121. IEEE,(2011)

[5] Umaparvathi, M., Varughese, D.K.: Evaluation of symmetric encryption algorithms for MANETs. In: International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1 – 3. IEEE , Coimbatore, (2010)

[6] Chang, J.T., Gundala, Moh, S., Moh, M.: VESS - a Versatile Extensible Security Suite for MANET Routing. In: Pacific Rim Conference on Communications, Computers and Signal Processing, pp. 944 – 950. IEEE, Victoria,(2009)

[7] Raj, N., Bharti, P., Thakur, S.: Vulnerabilities, Challenges and Threats in Securing Mobile Ad-hoc Network. In: Fifth International Conference on Communication Systems and Network Technologies, pp. 771 – 775. IEEE, (2015).

[8] Michiardi, P., Molva,R.: Ad Hoc Networks Security. J. IEEE PRESS, 329 - 354 (2004)

[9] Kumar, S., Pruthi, G., Yadav, A., Singla M.: Security protocols in MANETs. In: Second International Conference on Advanced Computing & Communication Technologies, pp 530 – 534. IEEE,(2012)

[10] Bird, Steven, Loper, E., Klein E.: Natural Language Processing with Python. O'Reilly Media Inc. (2009)