Information Leakage Prevention Using Public Key Encryption System and Fingerprint Augmented with Apriori Algorithm

Mudasiru Hammed *tundemuhammedy2k@yahoo.com*
Department of Computer Science
The Federal Polytechnic, Ilaro
Ogun State, Nigeria

Jumoke Soyemi *jumoke.soyemi@federalpolyilaro.edu.ng*
Department of Computer Science
The Federal Polytechnic, Ilaro
Ogun State, Nigeria

Abstract

The increase in the use of the internet around the world provided easier way of communication and information sharing that has led to the huge challenge of data leakage on the network. In an academic environment such as higher institutions of learning, the need to ensure that access to data and sensitive information are given to authorized users become imperative. However, this is not always the case as security bridges are often experienced. This study proposed a RSA public key encryption system and biometric fingerprint augmented with Apriori algorithm to prevent information leakages. The fingerprint verifies the identity of the owner of incoming message and the Apriori algorithm is used as the detection system instead of biometric that requires additional hardware for detecting fingerprint. This study developed a system based on the proposed algorithm. The developed system was tested on Federal Polytechnic, Ilaro local area network achieving a high level of security that prevents interception of valuable data by intruders or eavesdroppers. The system developed RSA public key encryption and fingerprint augmented with Apriori algorithm thus provided the required security mechanism that prevents information leakage in a public environment.

Keywords: Information, RSA Public Encryption Algorithms, Biometric Fingerprint, Apriori Algorithm.

1. INTRODUCTION

The increase in the use of the internet all over the world made internet-based services very popular thus making the sharing of information and communication much easier. This drastic rise in the use of network-based systems has rendered cyber security systems outdated because there is a rise in the activities of hackers and attackers using sophisticated methods such as phishing and spoofing. Therefore, the need to ascertain that authorized users and clients are the only one with access to both secure and sensitive information. The availability of tools such as the web tool, make credentials stealing and system cracking easier to achieve [1]. Data leakage is a frequent occurrence when dealing with confidential information detailing customer data, bank details, source code, design specifications and examination question paper. When such information leaks, the party concerned becomes insecure. Sharing data required an efficient asymmetric key encryption algorithm for preventing unauthorized user to gain access to vital information [2]. A number of research works in this area of study have been done to provide security in communication and sharing of data over network. These includes studies [3,4,5]. Tracing out data leakage among system users could prove very difficult for system administrator thus creating ethical issues in the working environment. Therefore, preventing information leakage is immensely advantageous than information leak detection.

This study proposed RSA public key encryption scheme that is more suitable for providing a security mechanism for tampering activities in exchanging information over the network and also incorporated fingerprint techniques augmented wih Apriori algorithm. This algorithm verifies the identity of an incoming message and the identity of the user sending the message which could be a panacea for spoofing attack. Apriori algorithm is a detection system that is embedded into administrator's server to verify the user generated fingerprints against fingerprints and details information obtained from the user during the registration.

2. LITERATURE REVIEW

There are various approaches for preventing sensitive data from unauthorized parties either to disclose or to modify communication between two parties. Among these approaches are cryptography, steganography, biometric, digital watermarking, digital signature and others. Cryptography and biometric however, are good at addressing threats, such as; information disclosure, tampering and spoofing. They form the foundation for all other aspects of information security.

2.1 Cryptography

Cryptographic schemes are grouped into symmetric (private key) cryptography and asymmetric cryptography [6]. In private encryption scheme, encryption must be equal to decryption key ($e = d$). However, malicious third party could work their way into intercepting the messages and gain access to sensitive information. Some could disclose or modify the information. Also eavesdropper are likely to be able to figure out the "secret key" except the secret key is securely exchanged [7]. It is obvious that RSA public key encryption technique is widely used because according to [8], there is a pair of keys, one which is known as the public key for the encryption of the plaintext that is meant to be sent and the other key known as private key. This private key is sent to the receiver for the purpose of message decryption.

2.2 Biometric

The biometric authentication works by scanning the user's characteristics such as finger print and eye retina and store information in the database in form of strings. Authentication of a user is achieved by matching the scanned data with that of the database with access granted based on the detection of commonalities [9]. However, Biometric authentication is good when limited applications are involved. Apart from the fact that the system slows down with large number of users, it also requires further hardware to detect the fingerprints and eye retinas. Fingerprints, however, can be spoofed simply by using common materials such as gelatin, silicone, and wood glue [10].

2.3 Apriori Algorithm

Apriori algorithm is the association analysis used to discover what is commonly called association rules. The algorithm looks at the possibilities of having items that are associated together in the transactional databases, with the support threshold and then identifies the recurrent item sets. The confidence threshold on the other hand uses restrictive probability that an item appear together in a transaction when another item appears, to locate association rules. This algorithm can be a potential strategy for detecting fake fingerprint without any additional hardware.

2.4 Related Work

Many research work proposed RSA public key encryption system for different purposes such as e-voting system, agent technology, electronic commerce and other sensitive data that needed to be prevented from unauthorized parties. Hybrid RSA Encryption Algorithm for Cloud Security was proposed by [4]. The study discovered that data decryption with hybrid RSA encryption algorithm is difficult to achieve when data transfer in the cloud system is applied. However, this hybrid encryption cannot ensure data transmission through secured protocol and it cannot also verify who the user is.

Study [2] proposed information leakage detection system using fingerprint data. This method generated fingerprint and token ID that are available in the database. The idea is to eliminate types of phrases from the fingerprinting process. Types of phrases are identified by looking at available public documents of the organization that is to be protected from information leakage and different phrases are identified with the help of databases. This process is only used for text file and if the text file contains some unicode characters in such cases, this process cannot give correct result.

In study [11], data encryption and decryption with RSA algorithm in a network environment was proposed. The algorithm allowed message senders to generate public keys to encrypt messages and the receivers' generated private keys using secured databases. An incorrect private key was able to decrypt the encrypted message but to a form that is different from the original message. An eavesdropper that breaks into the message will return a meaningless message despite the fact that the system cannot verify who the user is.

Study [12], developed an electronic voting system that used fingerprint technique. This technique took account of fake human thumb impression and separated out the authentic minutiae regions. The regions were subsequently extracted and used not only as user access control identity to the system but also to vote in the electoral process. The system developed was able to nullify multiple vote casting and provided election results collation process that is more accurate and secured authentication mechanism for voters. However, the cost of providing the special hardware device to implement biometrics enrolment and authentication was not considered. The possibility of providing the sensor with fake physical biometric could prove a security threat to the system.

Study [13] presented an online voting system using steganography and biometrics. The voters used their fingerprint which has been pre-enrolled to access the voting machine. The usage of anatomical traits rather than behavioral attributes further created more secured and acceptable voters registration system, because the fingerprint is unique, distinct, universal, and not easily damaged for every individual. Additional hardware is required to detect fingerprints. Fingerprints, in particular, can be easily spoofed from common materials.

In our study, the combination of RSA public encryption system and biometric fingerprint augmented with Apriori algorithm to prevent information leakages is proposed. This system ensures secured data transmission and also verifies who the user is and provides better way of securing data and information.

3. MATERIALS AND METHODS

This study focused on three processes which include: RSA key generation process, Encryption process, Validation and Decryption process.

3.1 RSA Key Generation Process

The use of encryption key requires every individual user to register with details information such as Name, Sex. Occupation, Level, Department, Marital status, E-mail, Phone number and their Fingerprint. The server stores the information and generate encryption key for the user in order to maintain user's account in the database. Figure 1 shows the encryption key generation process and Algorithm 1 shows RSA key generation described in [15] for the administrator to generate both public key and private key. The encryption key is made known to all users while the decryption key is kept secret to only administrator to decrypt any message sent by the users.

Algorithm 1: RSA Key Generation [14]

Administrator creates two different large and appropriately random primes, p_A and q_A which are kept secret.

Administrator calculates $m_A = p_A q_A$. The number known as the modulo is made public Administrator selects randomly some e_A, with $1 < e_A < (p_A - 1)(q_A - 1)$, making $\gcd(e_A, (p_A - 1)(q_A - 1)) = 1$. This number e_A is known as the encryption key which is made available to the user. Administrator calculates the inverse, $d_A = e_A^{-1}$ modulo m_A, of e_A. This number is kept secret. The pair (d_A, m_A) is Administrator's private key and d_A is called the decryption key. The private key will be used to decrypt any message received by the administrator. Administrator publishes the pair (e_A, m_A) as his public key

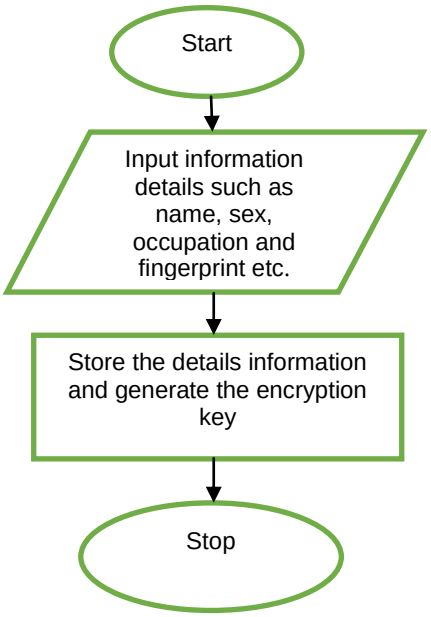

FIGURE 1: Encryption Key Generation Process.

3.2 RSA Encryption Process

For a user to send any sensitive information or document on the network, it has to be encrypted with the generated public key and fingerprint which verifies the identity of individual users. Once the information is encrypted, even the user can no longer decrypt the encrypted copy because there would not be private key to decrypt it. Algorithm 2 depicts the encryption algorithm.

Algorithm 2: Encryption Algorithm

Step 1: Obtained Administrator's public key (e_A, m_A)

Step 2: Take plaintext to be a positive integer M, where $0 < M < m_A$

Step 3: Calculate the cipher text $C = M^e \bmod m_A$

Step 4: Send the cipher text C to Administrator

The user use the algorithm 2 to convert plain text QBASICQUESTIONFORND to cipher text 12%&#345@!344!!@#27$*@24579#$$%&*#%

3.3 Validation Process

When an administrator receives the encrypted message like this; 12%&#345@!344!!@#27$*@24579#$$%&*#%, the identity of the sender has to be verified with the aid of Apriori algorithm by matching the sender fingerprint with the fingerprint already stored during the registration. The fingerprint extractor uses minutiae points and wrinkles features, any

extracted feature information is a template. This template is always comparing with newly presented extracted feature information using apriori algorithm. If commonality is not achieved, the system will automatically give error, but if commonality is achieved, the system will display user's information. Apriori **also** compares the template to all enrolled users to ensure that there is no similar template. If any similar template is found then the system will automatically give error and such user will never be activated. The algorithm 3 is Apriori algorithm described in [16].
Algorithm 3

Step 1: K=1
Step 2: $F_k = \{i \mid i \in I \wedge \mathrm{r}(\{i\}) \geq \mathrm{Nxminsup}\}$
{*Find all frequent 1-itemsets*}
Step 3: Repeat
Step 4: $K = k + 1$
Step 5: $C_k = apriori - gen\ (F_{k-1})$
{Generate Candidate itemsets}
Step 6: For each transaction $t \in T$ do
Step 7: C_t = subset (C_k, t) .{Identity all candidates that belong to t}
Step 8: For each candidate *itemset* $c \in C_t$ do
Step 9: $\sigma(\mathrm{c}) = \sigma(\mathrm{c}) + 1$. {Increment support count}
Step 10: *endfor*
Step 11: *endfor*
Step 12: $F_k = \{c \mid c \in c_k \wedge \sigma(\mathrm{c}) \geq Nx\mathrm{minsup}\}$
{Extract the frequent k-itemsets}
Step 13: until $F_k = \phi$
Step 14: Result = $\bigcup F_k$

Assuming the user's generated fingerprint is taken as X and the fingerprint obtained during the registration is Y. It means X and Y are conjunctions of attribute value-pairs, and s (for support) is the probability that X and Y appear together in a database and c (for confidence) is the conditional probability that X appears in a database when Y is present. The association rule X→Y is interpreted as data set that satisfies the conditions in X and also likely to satisfy the conditions in Y. This can be written as follows:

$s(X \longrightarrow Y)$ for support rule to make decision when finding relationship between two fingerprints

$c(X \longrightarrow Y)$ for confidence rule to measure the reliability of decision made

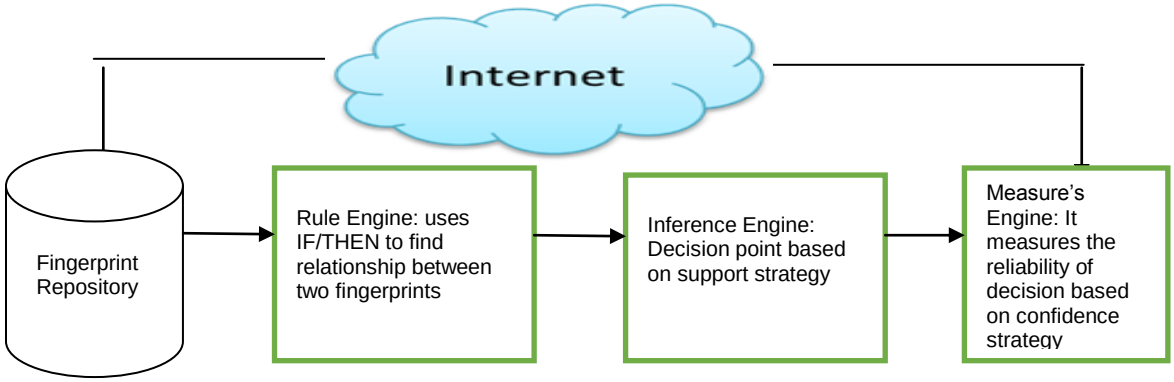

FIGURE 2: Architecture for Fingerprint Detection System.

The step 5 of algorithm 3 performs the following operations;
(i) Fingerprint Generation: this operation generates user's fingerprint and matches it with the already stored fingerprint.
(ii) Fingerprint pruning: This operation eliminates any fingerprint using the support-based pruning strategy

The flow flowchart in figure 3 depicts validation process to verify the identity of the incoming message before it is decrypted.

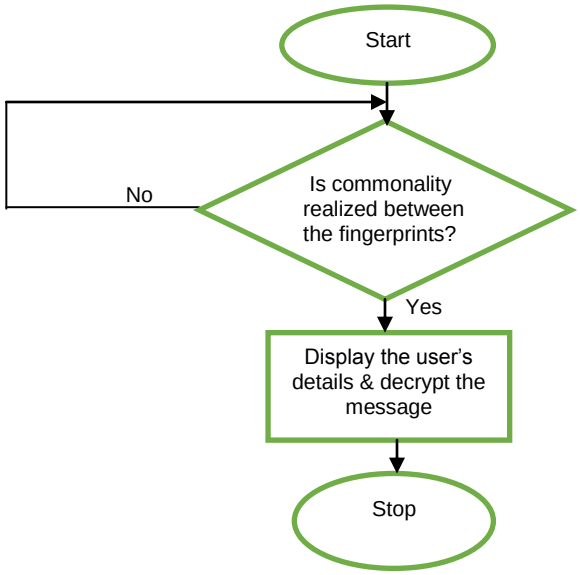

FIGURE 3: Flowchart of Validation Process.

3.4 RSA Decryption Process

After the identity of the message owner has been verified, the administrator will proceed to extract plain text Y from cipher text C, that is, 12%&#345@!344!!@#27$*@24579#$$%&*#% which will be converted back to QBASICQUESTIONFORND. Administrator follows steps in algorithm 4 to know the content of cipher text C.

Algorithm 4
Step 1: Obtained the cipher text C from user
Step 2: Administrator uses his private key (d_A, m_A) to compute $M = C^d \bmod m_A$
The two processes (encryption and decryption) allow both sender (administrator) and the receivers (lecturers) to communicate in a secured manner and the figure 4 depicts system architecture.

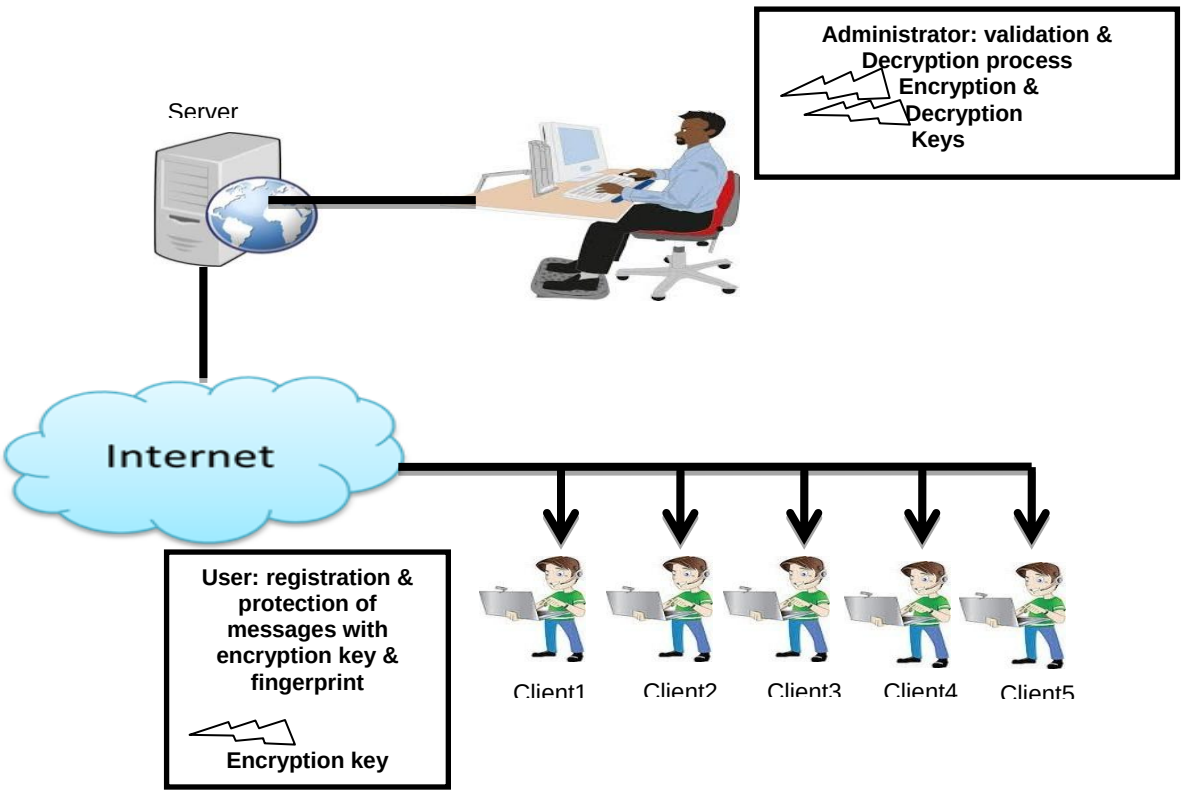

FIGURE 4: System Architecture.

4. RESULT AND DISCUSSION

Prevention of information leakage using public key encryption system and fingerprint augmented with Apriori algorithm was implemented on Federal polytechnic Ilaro, Ogun state local area network. PHP was used to design security application program and Jquery and MySQL were also used to allow communication between user and administrator. All users registered with their detail information such as name, sex, department, phone number, email address and fingerprint as it is shown in figure 5.

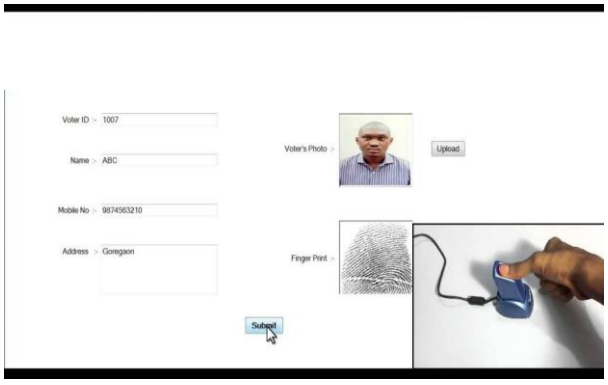

FIGURE 5: Biometric Fingerprint for Identification.

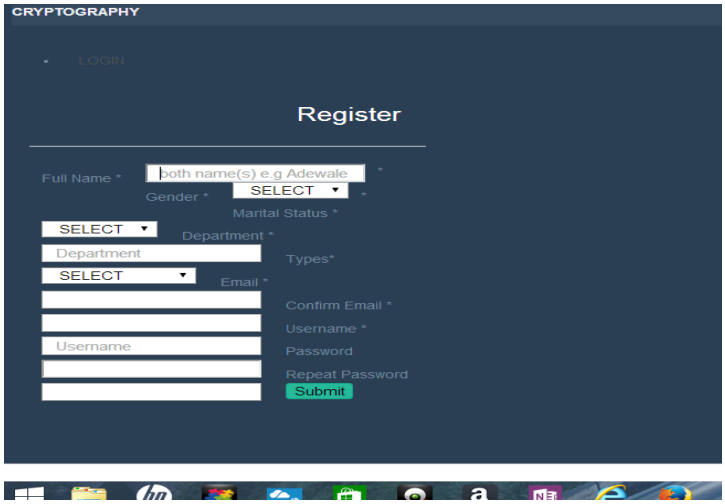

FIGURE 5.1: Registration Module.

The encryption key (e_A) is given to only successfully registered users as is shown in figure 6 while pair of that key (e_B) will be kept with the administrator.

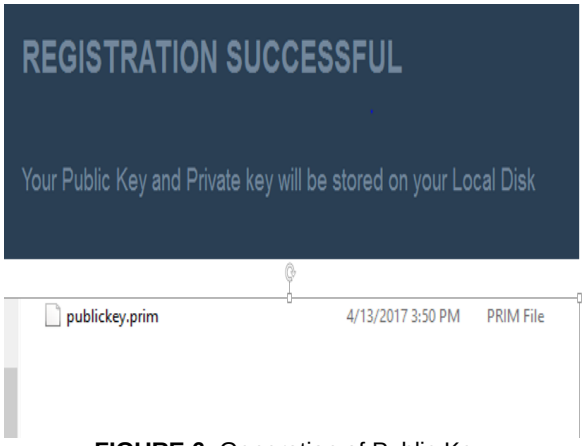

FIGURE 6: Generation of Public Key.

Subsequently, the users used the pair of encryption key (e_A) to protect sensitive information such as examination question papers, student results and other sensitive data before it is sent to the administrator in the server through the network. The encryption process ensures that sensitive data are delivered privately and makes it so hard for intruders to bypass as it is shown in figure 7.

FIGURE 7: Encryption Process.

The fingerprint used in the study allows the administrator in the server to verify all the received messages sent to the server to know whether it comes from the right user (registered user). The Apriori algorithm was implemented to verify the sender's fingerprint, if the sender is one of the registered users in the system, it will automatically display the user's information. After the identity of the sender has been confirmed, the administrator will use its private key (d_A) to decrypt all the received messages as it shown in figure 8 and 9. But any sent message that its source cannot be verified (does not come from the registered user) will not be decrypted. The fingerprint augmented with Apriori algorithm ensured that identity of any message must be verified.

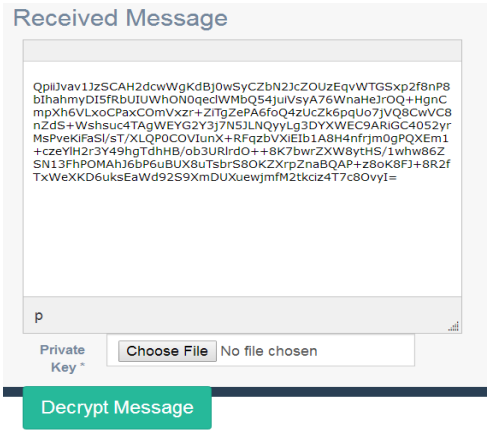

FIGURE 8: Cipher-text Message Received from Administrator.

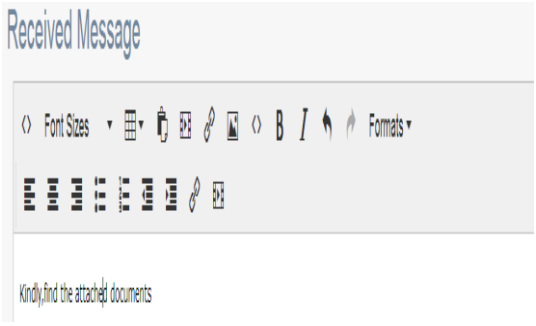

FIGURE 9: Decryption Messages.

The validity of RSA public encryption system and biometric fingerprint augmented with Apriori algorithm for preventing information leakages was tested using institutional local area network. One of the threats tried during testing include but not limited to messages sent by unregistered users and the messages were neither decrypted nor lost but were all privately delivered. The developed system achieved high degree of security making it difficult for intruders or eavesdroppers to intercept valuable data such as examination question and the students' results. Public key encryption system and fingerprint augmented with Apriori algorithm improves security mechanism for preventing information leakage in an open network. Apriori algorithm was used as a fingerprint detection technique instead of biometric fingerprint techniques that requires additional hardware to detect fingerprint.

5. CONCLUSION

Public encryption key and Fingerprint augmented with Apriori algorithm used here, guarantees enough security mechanism for preventing information leakages and unauthorized users. The students and any unauthorized staff cannot gain access to sensitive information such as examination question papers and student examination results before formal release of the results. These two techniques gained a high level of security mechanism for securing information with high degree of accuracy and reliability. The algorithms employed in the work enhanced the performance and security aspect of the system. Several studies have used public key encryption system to secure information and data, but to verify the identity of the message before it is decrypted has always been a great challenge. This study used biometric fingerprint to verify the identity of the message before it is decrypted, with the augmented Apriori algorithm overcoming the biometric fingerprint short comings by removing additional hardware for information analysis which gets quite slow with increase in population.

FUTURE WORK

The system developed in this study was tested with the Institution's local area network which is a closed network. There is need to increase the scalability of the system by developing a system that works with open network.

6. REFERENCES

[1] H. Dinne, K. Mandava. *Two Way Mobile Authentication System*. MA Blekinge Institute of Technology: Karlskrona, Sweden. 2010.

[2] R. Sinha, C. Choudhary. "Information leak detection system using fingerprint of data", *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, no. 12, pp. 3911-, 2014.

[3] A. Patel, R. Kansara and P. Virpari. "A Novel Architecture for Intrusion Detection in Mobile Ad hoc Network". *International Journal of Advanced Computer Science and Applications*, pp. 68-71.

[4] S. Nandita. "Designing of Hybrid RSA Encryption Algorithm for Cloud Security". *International Journal of Innovative Research in Computer and Communication Engineering,* vol. 3, no. 5, pp. 4146-4152, 2015.

[5] L. Paul, M.N. Anilkumar. "Authentication for Online Voting Using Steganography and Biometrics". *International Journal of Advanced Research in Computer Engineering & Technology*, vol. 5, no. 10, pp. *26*–32, 2012.

[6] A. Singh, R. Gilhotra. "Data security using private key encryption system based on arithmetic coding". *International Journal of Network Security & Its Applications (IJNSA)*, vol. 3, no. 3, pp. 58-67, 2011.

[7] S.T. Vuong and P. Fu A security architecture and design for mobile intelligent agent systems. *ACM SIGAPP Applied Computing Review*, vol. 9, no. 3, pp. 21-30, 2001.

[8] H.K. Al-Anie, M.A. Alia and A.A. Hnaif. "eVoting Protocol Based on Public-Key Cryptography". *International Journal of Network Security & Its Applications (IJNSA)*, vol. 3, no. 4, pp. 87-98, 2011.

[9] H.B. Kekre and V.A. Bharadi. "Using Component Model for Interfacing Biometric Sensors to Capture Multidimensional Feature". *International Journal of Intelligent Information Technology Application*, vol. 2, no. 6, pp. 279-285, 2009.

[10] A. Wiehe, T. Søndrol, O.K. Olsen and F. Skarderud. *Attacking fingerprint sensors*. *Gjøvik University College,* 2004.

[11] N.Y. Goshwe. "Data encryption and decryption using RSA algorithm in a network environment". *International Journal of Computer Science and Network Security (IJCSNS)*, vol.13, no. 7, pp. 9-20, 2013.

[12] S. Kumar and M. Singh. "Design a secure electronic voting system using fingerprint technique". *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 4, pp.192-202, 2013.

[13] N. Malwade, C. Patil, S. Chavan and S.Y. Raut. "Secure Online Voting System Proposed by Biometrics and Steganography". *International Journal of Emerging Technology and Advanced Engineering*, 2013.

[14] Jean H G. Notes on RSA. *University of Pennsylvania Scholarly Commons.* 2010, *pp. 329-347.*

[15] E.H. Han, G. Karypis and V. Kumar. *Min-apriori*: *An algorithm for finding association rules in data with continuous attributes*. Department of Computer Science and Engineering, University of Minnesota, Tech. Rep. 1997.