International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

A Novel Model for S.M.S Security and SPAM Detection

Nikhila Zalpuri[1]

[1]Dr. APJ Abdul Kalam Technical University, Computer Science and Engineering, Sector-62, Noida, India

Abstract: ***Short message service (S.M.S) security is a very important issue in today's time. In spite of the fact that instant messengers have become very popular, S.M.S is still used. S.M.S is prone to several attacks and is not yet marked as a safe means for communication. Security in S.M.S includes message authentication, user authentication, encryption and decryption of the message and Spam detection .This paper provides a model which provides security for all these aspects of S.M.S. It is found that when this model is implemented the accuracy of is 94%.***

Keywords: S.M.S, spam, ElGamal, ann, SVM, HMAC

1. Introduction

With the increasing number of cell phone users, the use of S.M.S is increased. S.M.S is not only used by users but also by many organizations like the bank, military, etc to exchange useful information. This information need to be secured in order to prevent the loss of people, organizations or the country. Recent research performed by Acision [16] revealed that SMS is still the most prevalent messaging service in the U.S. when compared to other services. 61% of respondents said they own a smartphone today, with 91% of this demographic reportedly using SMS regularly even with various other instant messaging services on their handsets. 65% of these SMS users said they need the service today and 45% of them even said they would be lost without it. Thus the development of the security system for SMS is very important.

According to the recent researches done by Kaspersky Laboratory (2014), almost 65.7% of all emails were considered as spam, respectively in January. In this regards, a huge amount of bandwidth is wasted and an overflow occurs while sending the emails. In fact, separating spam from legitimate emails can be considered as a kind of text classification, because the form of all emails is generally textual and by receiving the spam, the type has to be defined. Support Vector Machines are supervised learning models or out-performed other with associated learning algorithms and good generalization that analyze data and recognize patterns, used for classification and regression analysis [15]. According to reported statistics United State of America, China and South Korea are among the main sources of these spam respectively with 21.9%, 16.0% and 12.5%. Fig. 1 shows the spam sources for each country [14].Indicating that Asia is the biggest spam producer out of all the countries and hence it can be said that it is very important to enforce spam filtering in countries like Asia. In this paper, we have surveyed all the above given techniques so as to create an architecture for S.M.S exchange that provide total protection.

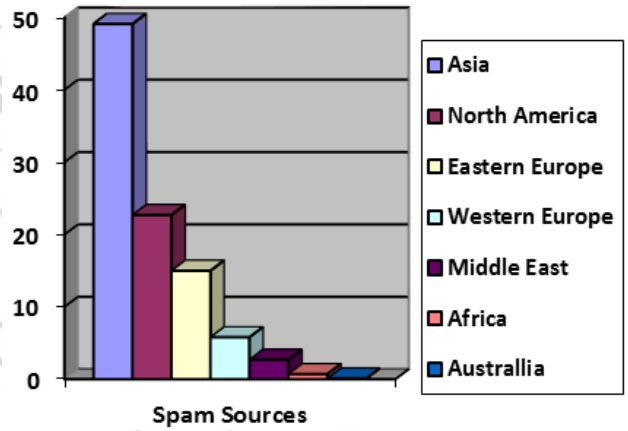

Figure 1: Spam Sources -Country wise

This paper provides a model which takes into consideration all the security issues in a SMS. This is all done by using HMAC-Hash Message Authentication Code i.e. SHA-1, SHA-256, SHA-384, SHA-512, MD-5 and MD-2 for user authentication, ElGamal for encryption and decryption, RSA with MD-5 for Digital signature, and SVM-Support Vector Machine to check if the SMS is spam or not.

To develop this security system we need to first understand Hash functions, ElGamal, RSA, and SVM.

Firstly, hash functions are common and critical cryptographic primitives. Their primary application is combined use with public-key cryptosystems in digital signature schemes. The most widespread functions are SHA-1 (Secure Hash Algorithm- 1) and MD5 (Message Digest). These two hash functions are widely known for being used in the Keyed-Hash Message Authentication Code (HMAC), which is met in numerous communication applications, to address authentication issues.

Secondly, the ElGamal system is a public-key cryptosystem based on the discrete logarithm problem. It consists of both encryption and signature algorithms. The encryption algorithm is similar in nature to the Diffie-Hellman key agreement protocol The ElGamal signature algorithm is similar to the encryption algorithm in that the public key and private key have the same form, however, encryption is not

the same as signature verification, nor is decryption the same as signature creation as in RSA. DSA is based in part on the ElGamal signature algorithm. Analysis based on the best available algorithms for both factoring and discrete logarithms shows that RSA and ElGamal have similar security for equivalent key lengths [17].

Thirdly, the RSA (Rivest-Shamir-Adleman) signature scheme is a deterministic digital signature scheme which facilitates message verification and recovery. Digital signature plays an important role in online communication. In these days most of the electronic documents are identified by the digital signature only. Message digest algorithm is used to generate message digest of a given input message. Message digest is also called hash code or finger print of the input message. [1].

Lastly, Support vector machines (SVMs) are relatively new approach that has rapidly gained popularity because of the very fast and accurate results they have achieved in a wide variety of machine learning problems. Support vector machine algorithms divide the n dimensional space represented data into two regions using a hyper plane. This hyper plane always maximizes the margin between the two regions (classes). The margin is defined by the longest distance between the examples of the two classes and is computed based on the distance between the closest instances of both classes to the margin, which are called supporting vectors. Support Vector Machines is also a supervised learning method for automatic pattern recognition. SVM learns from a training data set to classifier which separates a set of positive examples from a set of negative examples of introducing the maximum margin between the two data sets. The training data set can be described by points in the dimensional space [2].

This paper deals with the survey of all these aspects of security in order find the best options available for each of them today .It is found that for message authentication many hash algorithms are available and out of those the most widespread functions are SHA-1 (Secure Hash Algorithm-1) and MD5(Message Digest). It was also found that for message encryption and decryption of S.M.S the best available algorithm is ElGamal algorithm which produces shorter keys which are more secure and SVM (Support Vector Machine) is the best machine learning technique for SPAM detection. All the above stated techniques are combined together to make one single model for ultimate protection.

The advantage that this model has over the available security models is that this model is considering the best available security techniques available for message authentication, user authentication, encryption/ decryption and spam detection classifier by comparing the techniques with the one‟s available in today‟s market and combining them together.

The rest of the paper is organized as follows. Section II presents the literature survey of the papers studied. In Section III lists the threats to which SMS communication is prone. Section IV presents the model for SMS security in detail. In Section V the overview of the algorithms studied during the course of this review. Section VI shows the results and analysis. Finally Section VII summarizes the conclusions.

2. Literature Review

Marko Hassinen [4] has used RSA algorithm to encrypt SMS messages used in mobile commerce, whereas keys are generated using SHA-1. Private keys are restricted to mobile devices. Authentication Server will then generate certificates for public keys. Light weight Directory Access Protocol (LDAP) database is used to store/retrieve those certificates. These certificates are further used by mobile user to exchange encrypted SMS messages.

Myungsun Kim [3] used El-Gamal encryption scheme to decompose extension fields. It is also used to decompose the public key using El-Gamal Encryption method. It helps to reduce multiple cipher-texts without losing any information. El-Gamal encryption scheme consists of Key generation, Encryption, and decryption. Private and public keys are generated in Key-gen step. Encoding and encryption of plaintext is done in next step. Finally, shared secret key is defined for receiver and message is decrypted using that shared key.

Er. Kumar Saurabh [5] proposed a new method for nodes authentication in wireless sensor networks using RSA. RSA algorithm is applied into source node, intermediate node, and destination node. Proposed algorithm generates private and public keys. Then cipher text is created, which is encrypted using public key. Private Key is sent to the receiver. After encryption, packet is sent to intermediate node, which sends it to the destination node. Destination node will finally decrypt it using private key.

David Lisoněk [6] proposed an application to encrypt SMS messages using asymmetric RSA cipher. OAEP padding scheme is used to avoid RSA from dictionary attacks. Private keys are stored in the application, whereas public keys are stored in mobiles memory. Symbian OS is used as a programming environment since it requires less computational power. Key generation operation is tested on Nokia N80 by subtracting the actual start time of key generation from its final time. Analysis of several attacks on application is also conducted at the end.

Alfredo De Santis [7] proposed a secure extensible and efficient SMS (SEESMS) application framework which allows two mobile peers to exchange encrypted SMS message in an efficient manner by selecting their level of security. ECIES and RSA are used for encryption. RSA, DSA, and ECDSA signatures are also used to validate contacts. After being registered with SEESMS on mobile, keys are exchanged between users to transmit secure SMS using HMAC. Users will then select energy efficient cryptosystem, encrypt SMS using it, and send to the receiver. Comparison of RSA, DSA, and ECDSA is conducted on the basis of energy efficiency on N95 mobile. RSA and DSA are found better than ECDSA.

Neetesh Saxena [8] has analyzed and compares different digital signature methods, i.e., DSA, RSA, and ECDSA using Java. The experiments are conducted on PC to check encryption performance of all three algorithms. Results of RSA, DSA, and ECDSA are shown on the basis of their key generation execution time, signature generation, and signature verification time. It is found that SHA-1 provides better security and ECDSA is better than DSA in signature generation and verification. Results have shown that proposed ECDSA performs better than simple ECDSA.

Ioannis Yiakoumis, Markos Papadonikolakis, Harris Michail and Athanasios P.Kakarountas [9] presented a design approach to create small-sized high speed implementations of Keyed-Hash Message Authentication Code (HMAC) .The proposed implementation can either operate in HMAC-MD5 and/or in HMAC-SHA1 mode. The proposed implementations do not introduce significant area penalty. However the achieved throughput presents an increase compared to commercially available IP cores that range from 30%-390%. The main contribution of the paper was the increase of the HMAC throughput to the required level to be used in modern telecommunication applications, such as VPN and the oncoming 802.11n.

Vinod Patidar [10] worked on E-mail spam classification. Among the approaches developed to stop spam emails, filtering is a popular and important one. Common uses for email filters include organizing incoming email and computer viruses and removal of spam. As spammers periodically find new ways to bypass spam filters and distribute spam messages, researchers need to stay on the forefront of this problem to help reduce the amount of spam messages. Currently spam emails occupy more than 70% of all email traffic. The negative effect spam has on companies is greatly related to financial aspects and the productivity of employees in the workplace. This paper proposes the new approach to classify spam emails using support vector machine. It found that the SVM outperformed than other classifiers.

Venkateswara Rao Pallipamu, Thammi Reddy K, Suresh Varma P [11], presented a paper on Digital Signature. It is used to achieve non-repudiation service, which provides proof for sent or received messages. In this paper they proposed a new digital signature scheme using a novel message digest algorithm, „Algorithm for Secure Hashing-160 (ASH-160)". This proposed scheme was implemented in java and the results were analyzed and compared with RSA digital signature scheme using SHA1 and RIPEMD160. The analysis of experimental results reveals an increase in security strength and slight improvement in the efficiency of RSA with ASH160 than the compared schemes.

3. SMS Security Threats

A. Message Disclosure

In the SMS service message is transmitted as an unencrypted text. Message could be intercepted during transmission. SMS is first stored as an unencrypted text in the SMSC and then delivered to the destination receiver. This message could be viewed by the users in the SMSC. ElGamal encryption approach, secure the transmitted SMS from Message Disclosure attack.

B. Man-in-the-middle Attack

When the user does not authenticate then th attacker can either read and alter the transmitted SMS or send the SMS using someone else"s name. To check if the man-in-the-middle has modified the message or not Hash Message Authentication Code can be used (HMAC). To check if the person sending is authentic or not RSA algorithm along with message digest can be used. This approach will save from man-in-the-middle attack

C. SMS Viruses

There have been no reports of viruses with message when the message is transfer from one mobile device to another but mobile devices are getting more powerful and programmable. The SMS viruses are being spread through the message. Generally these viruses containing SMS are spam SMS, so if detection of the spam SMS is possible then this can save smartphones from viruses [12].

D. SMS Spamming

While using SMS as a legitimate marketing channel, many people have had the inconvenience of receiving SMS spam. The availability of bulk SMS broadcasting utilities makes it easy for virtually everyone to send out mass SMS messages. Detection of spam SMS will solve this problem.

4. Methodology

This model is a very secure one as all the aspects of security used while sending and receiving a S.M.S are considered. Spam detection is done using SVM classifier and the best available options for security are combined together to provide ultimate protection.

First user is asked to enter the user number (the user number is considered to be phone number) then he is asked to enter the message, the message to be sent by sender and the message received by the receiver. During the sending-receiving phase encryption and decryption is performed. After receiving the message the message is classified as spam or not spam.

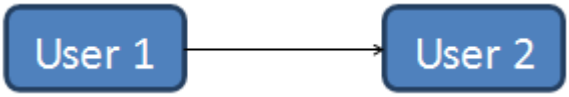

Figure 2: User 1 Send a Message to User 2

The four security issues is handled by these methods. First the authenticity of the user is checked by using RSA-Algorithm along with MD-5, only the senders marked by the receiver is considered authentic rest all is considered as unauthentic. The different values of p and q is assigned to each user, and then user is classified to prevent spoofing attack. Then ElGamal technique is used to encrypt the sent message and decrypt the received message. This is done to provide.

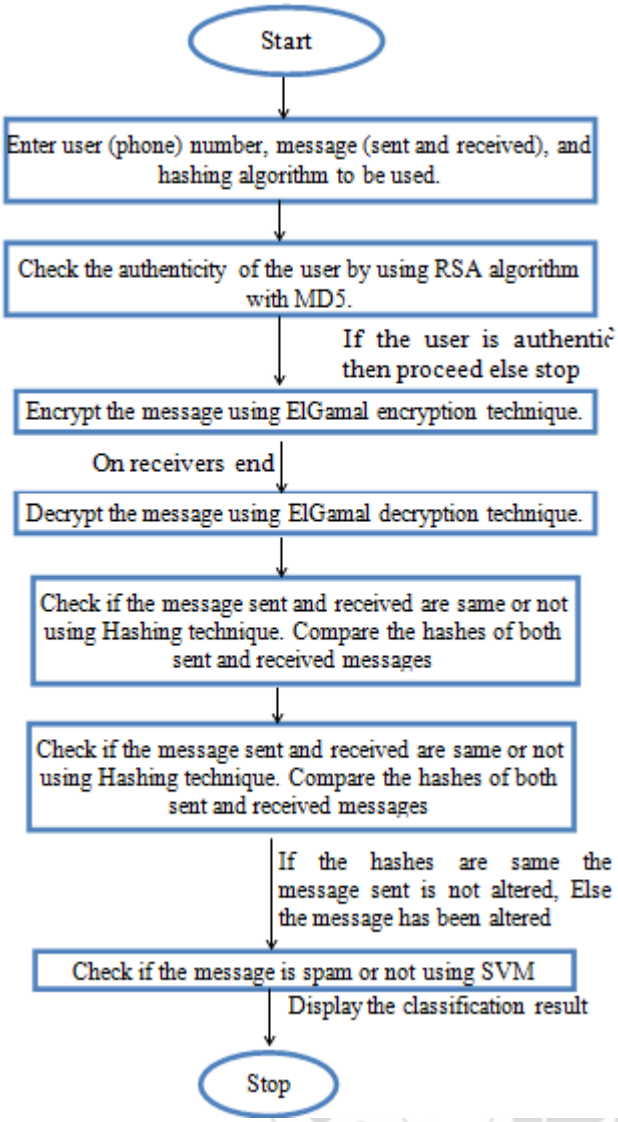

Figure 3: Flow chart of the Model

After this the sent message and received message is used to check if they are same or not. To check this secure hashing algorithm and message digest algorithm, one at a time, SHA-1, SHA-256, SHA-384, SHA-512, MD-2 and MD-5is used. This depends on user that which one should be used when according to requirements. This is done to check if the integrity of the message is maintained or not, i.e., if the message wasn't altered during transmit.

Last part is done using the received message to check if the message was a spam message or a non-spam message, if the received message is classified as spam message then the user can delete the message without opening that particular confidentiality, to save the SMS from being read by any unauthorized person in the middle, Man-in-the-middle attack message to save his phone from the SMS viruses. The classification is done using support vector machine.

All these techniques are combined in a single model to provide great SMS security under one roof.
Pseudo code for the model:
Step 1: Start.
Step 2: Input the message the sender wants to send and the receiver receives and phone number of the sender.
Step 3: Calculate hash for both the message received and sent.
Step 4: If hash is equal, message is declared as untampered and authentic:

Hash1=Hash2 (1)

Step 5: Encrypt the sent message on the receivers end using ElGamal algorithm.
Step 6: Decrypt the message at receivers end.
Step 7: Check if user is authentic or not using his phone number.
Step 8: Detect if the message is SPAM using SVM classifier.
Step 9: Display results.
Step 10: Stop

The architecture of the model is as shown in figure 2 and 3.Whenever User1 wants to communicate with user 2 in figure 2, he will have to follow the steps mentioned in figure 3, as user1 will follow all steps mentioned from message authentication, encryption and decryption of the message, user authentication and Spam detection, this model will provide complete security under one roof.

5. Overview of the Algorithms Used

A. Secure Hash Algorithm

The Secure Hash Algorithm is a family of cryptographic hash functions published by the National Institute of Standards and Technology (NIST) as a U.S. Federal Information Processing Standard (FIPS), including:

- SHA0: A retronym applied to the original version of the 160bit hash function published in 1993 under the name "SHA". It was withdrawn shortly after publication due to an undisclosed "significant flaw" and replaced by the slightly revised version SHA1.
- SHA1: A 160bit hash function which resembles the earlier MD5 algorithm. This was designed by the National Security Agency (NSA) to be part of the Digital Signature Algorithm. Cryptographic weaknesses were discovered in SHA1, and the standard was no longer approved for most cryptographic uses after 2010.
- SHA2: A family of two similar hash functions, with different block sizes, known as SHA256 and SHA512.

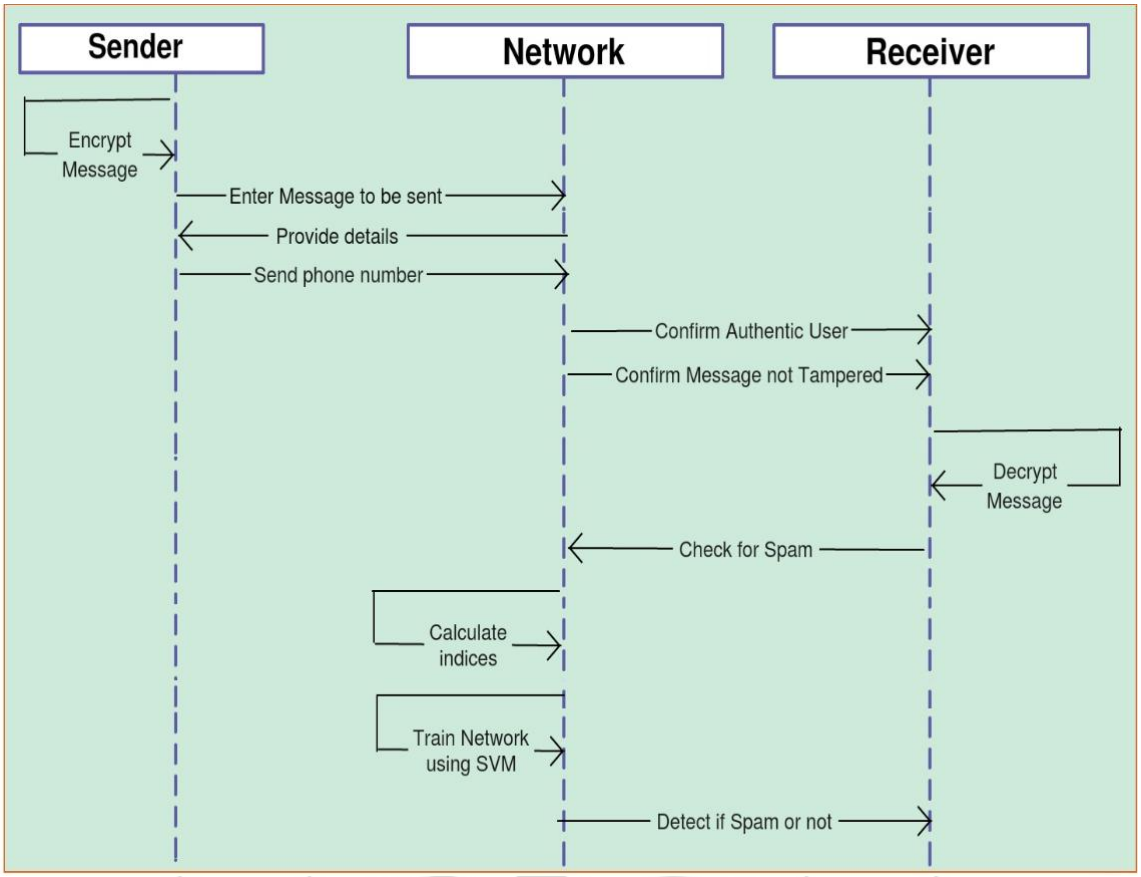

Figure 4: Sequence diagram of the Model

They differ in the word size; SHA256 uses 32bit words where SHA512 uses 64bit words. There are also truncated versions of each standard, known as SHA224, SHA384, SHA512/224 and SHA512/256. These were also designed by the NSA.

B. ElGamal

ElGamal public-key encryption is a continuation of Diffie-Hellman key exchange protocol. This system was defined by Taher Elgamal in 1984. There are three main components which are the key generated; the encryption and decryption algorithms. ElGamal is characterized as asymmetric cryptography and it is commonly known that the encryption and decryption speed is slower than in a symmetrical algorithm [17].

C. Rivest Shamir Adleman Algorithm (RSA)

Rivest Shamir Adleman Algorithm is one of the most challenging algorithms. This algorithm is develop by the Ron Rivest, Adi Shamir, and Len Adleman in 1997.The RSA algorithm is one of the most widely used algorithm and the implementation of this algorithm is very simple as compared to the another algorithm. It consists of the encryption and decryption to encrypt and decrypt the data. The key size of the RSA algorithm is larger than the Elliptic Curve Cryptography. The RSA algorithm consists of the prime number and the product of the prime number forms the encryption key. This encryption is used to secure the data in the system. In RSA algorithm consist of many operations. One of the most important operations is the modular exponentiation. By using this operation we can encrypt and decrypt the message. Many attacks are bombarded on the RSA algorithm and these attacks are hold successful against the RSA algorithm. RSA algorithm is used to encrypt and decrypt the data on both side i.e. sending and receiving ends. In this Algorithm plaintext is encrypted in the block. RSA algorithm consist an integer and the binary values. In the RSA algorithm message is encrypted using the asymmetric encryption cipher. RSA algorithm consists of the large exponent and the efficiency is also large [13].

6. Results and Analysis

The RSA technique was considered the best as compared by Alfredo De Santis, Venkateswara Rao Pallipamu, Thammi Reddy K, Suresh Varma P, and Er. Kumar Saurabh thus it is used along with the simplest and most effective and commonly used message digest MD-5.

The ElGamal technique is used because it is an asymmetric cryptographic algorithm and provides better result than other algorithm as discussed by Myungsun Kim[3].

All the hashing techniques and message digest techniques are used as all gave good results as compared by Ioannis Yiakoumis, Markos Papadonikolakis, Harris Michail and Athanasios P. Kakarountas[9]. So we have used all of them. This will also give user a set of choices to choose different hashing techniques for different purpose and security level.

Support vector machines were considered best by Vinod Patidar[10] for E-mail spam classification, thus this technique is used in SMS spam classification.

The Accuracy of the spam detection in S.M.S is calculated using SVMlib code:

A.Prepare dataset

```
iris = load_iris()
X = iris.data[:, :2]
y = iris.target
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2)
```

B.SVM classification

```
clf = svm.SVC(kernel='rbf', gamma=0.7,
C = 1.0).fit(X_train, y_train)
y_predicted = clf.predict(X_test)
```

C.Performance

```
print "Classification report for %s" % clf
print
print metrics.classification_report(y_test,y_predicted)
print
```

Which will produce the output as shown in Table 1:

Table 1: Output of the Accuracy Test

	Precision	Recall	F1-Score
0	1.00	1.00	1.00
1	0.96	0.69	0.90
2	0.80	0.91	0.98
Average	0.88	0.80	0.94

Abbreviation: tp=true positive, fp= flase positive, tn=true negative, fn= false negative

$$\text{Precision} = \frac{tp}{tp + fp} \quad (2)$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (3)$$

The Accuracy (F1 score) is 94%.

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

The Line chart for the Precision and recall is shown in Fig. 5

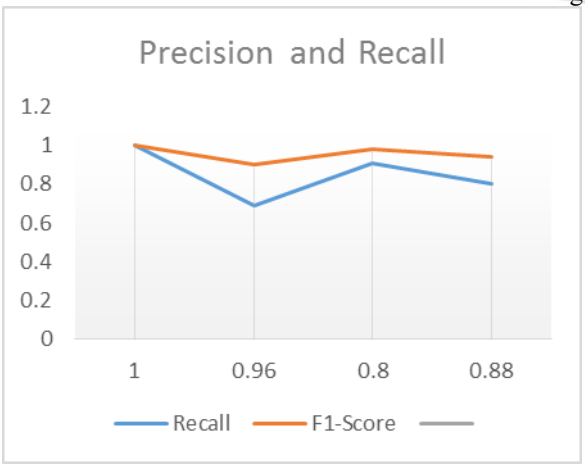

Figure 5: Accuracy Chart

7. Conclusion

This paper provides a model which combines all the best possible techniques available to secure SMS and has the capability of providing complete security and includes message authentication, encryption/ decryption of the message, user authentication and spam detection. On the basis of the results obtained in the papers studied, we can say that the collection of all the best techniques under one roof will provide ultimate SMS security which is MD5 with rsa for user authentication, SVM for spam detection and Elgamal for encryption/decryption of the message. The accuracy for the model comes out to be 94%.

References

[1] S.Sharma, J.S.Yadav and P.Sharma, "Modified RSA Public Key Cryptosystem Using Short Range Natural Number Algorithm",IJARCSSE, volume 2, August, 2012.

[2] S.Amari, S.Wu,"Improving support vector machine classifiers by modifying kernel functions". Neural Networks, Page No (783– 789),1999.

[3] M. Kim, J. Kim, and J. H. Cheon, "Compress Multiple Ciphertexts using ElGamal Encryption Schemes", J. Korean Math. Soc, vol. 50, Page No (361-377),2013.

[4] M. Hassinen, "Java based public key infrastructure for sms messaging", Proc. 2nd International Conference on Information and Communication Technologies,ICTTA'06, Page No (88-93), 2006.

[5] S. Singh and E. K. Saurabh, "Providing Security in Data Aggregation using RSA algorithm", International Journal of Computers & Technology, vol. 3, Page No(60-65) ,2012,.

[6] D. Lisonek and M. Drahansky, "Sms encryption for mobile communication", Proc. International Conference on Security Technology, SECTECH'08 ,Page No (198-201), 2008,

[7] A. De Santis, A. Castiglione, G. Cattaneo, M. Cembalo, F. Petagna, and U. F. Petrillo, "An extensible framework for efficient secure SMS", Proc. International Conference on Complex, Intelligent and Software Intensive Systems (CISIS) , Page No (843-850), 2010.

[8] N.Saxena and N. S. Chaudhari, "Secure encryption with digital signature approach for Short Message Service", Proc. World Congress on Information and Communication Technologies (WICT) , Page No (803-806), 2012.

[9] I. Yiakoumis, M. Papadonikolakis, H. Michail and A. P. Kakarountas, "Efficient Small-Sized Implementation of the Keyed-Hash Message Authentication Code", eurocon

[10] A. De Santis, A.Castiglione, G.Cattaneo, M.Cembalo ,F.Petagna, and U.F.Petrillo, " An extensible framework for efficient secure SMS", Proc. International Conference on Complex Intelligent and Software Intensive System,2010

[11] "Spam", http://www.kaspersky.com/about/news/spam, 2015

[12] Zahra S.Torabi, Mohammad H.NadimiShahraki, Isfahan,Iran and Akbar Nabiollahi,"Efficient Support Vector Machines for SpamDetection: A

Survey",(IJCSIS) International Journal of Computer Science and Information Security,Vol. 13, No. 1, January 2015.
[13] "Spam News", https://en.wikipedia.org/wiki/Acision,2015
[14] Bharat Gupta, Dr. Rajeev Gupta,"Performance Evaluation of Modified Signcryption Scheme", International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397, Vol. 4 No. 12, Page No (1881-1889), Dec 2012.

Author Profile

Nikhila Zalpuri was born in Noida India. She obtained B.E degree at the B.M.S College of Engineering, Bangalore and is persuing her M.tech degree in Computer Science (2013-2015) from J.S.S Academy of Technical Education, Noida at Dr. A.P.J Abdul Kalam Technical University,Lucknow.