IJCSE International Journal of Computer Science and Engineering Open Access
Research Paper Volume-2, Issue-3 E-ISSN: 2347-2693

Intrusion Detection and Violation of Compliance by Monitoring the Network

R. Shenbaga Priya[1*], V.Anusha[2] and N.Kumar[3]

[1*, 2, 3] *Department Of Computer Science with specialization in Network, Vel Tech MultiTech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, India*
shenbagapriyar@gmail.com; anushavijayakumar26@gmail.com; nkvsc.org@gmail.com

www.ijcseonline.org

Received: 02 Mar 2014 Revised: 10 Mar 2014 Accepted: 22 Mar 2014 Published: 31 Mar 2014

Abstract- Network and security of system has vital role in data communication environment. Web services and networks can be crashed on attempting many possible ways on forwarding by hackers or intruders. It causes malicious rapt in which it needs a technique called Intrusion Detection System through Spam Filtering. Thus gives the protection to networks. It can be done by using Open Source Network Intrusion Detection System called Snort. The process of arranging the e-mail with framed criteria called Spam Filtering. Proposed System, a Machine Learning Algorithm called Simple Probabilistic Navie Bayes Classifier used to detect the intrusion. Based on its content Probability of Spam messages can be calculated in Navie Bayes Classifier by learning it from spam and Good mail which results a robust, efficient anti-spam approach and adaption. Sniffing the packet and Fed it as input to Navie Bayes Classifier will give Test Dataset. Depends on spam and intrusion probability, the email is been classified as good or spam.

Key Terms- Intrusion Detection, Navie Bayes Algorithm, Spam Filtering, Dynamic Tuning Mechanism

1. INTRODUCTION

With the greater vogue of e-mail, many users and companies found it is an easiest way to forward a bulk amount of gratuitous messages to immense number of users at a low cost way. These uninvited bulk messages or junk e-mails are called spam messages [10]. The greater part of spam messages that has been reported recently are unsolicited commercials promoting services and products including sexual enhancers, cheap drugs and herbal supplements, health insurance, travel tickets, hotel reservations, and software products[4][10]. It can also comprises of abusive content such as obscene images and can be used as well for spreading gossips and other fraudulent advertisements such as make money fast[10]. Apart from many activities attackers are using spam e-mail to hack the system and acquire the details of the end users system[6]. E-mail spam has growing at a faster rate over the past couple of years[7]. It has become a paramount threat for commercial users, network administrators and even normal end users[7].

A study in July 1997 reported that, Spam can be very expensive to e-mail recipients. It diminishes their productivity by consuming their time and causing displeasure to deal with a huge amount of spam [3][4].

Objectives- It is paramount to set the aim of the paper before starting over the paper which needs to be realistic and achievable in a reasonable time frame. The goal of the paper is to frame a system that captures e-mail and examine it by detecting the Intrusion or violation of compliance using probabilistic Naive Bayes Classifier algorithm[5][1].

2. SYSTEM ANALYSIS

*Problem-*On internet everyday text files are available exponentially which increases the information of online[9]. Text Classification is a major component in information management tasks, as the assignment of text files are based on information contained in one or more pre-defined categories [7]. The fundamental concept of Naive Bayes used[1].Text Classifier are commenced for the design and implementation of the spam e-mail detector[4][5].

*Existing System-*Due to the enormous increase in the spam capacity, spam filtering has observed over the past few years. Various solutions comprise of business and products is been proposed and deployed [1].

White list is a technique called rule-based filtering , where a whitelist is a register consist of a collection of contacts from which e-mail messages can be received[6]. If an e-mail reaches from the contact which is not present in the whitelist, then it is considered as spam and put into spam folder[4][6]. This technique is effective for some extent; it has short come also [3]. Any email sent by a stranger will simply be considered as false positive (FP). However there is a strategy that includes a challenge response mechanism to allow users to be added to a user's whitelist[1].

A blacklist incorporate lists of known spammers [6]. Significantly when a user receives spam, then the user can

Corresponding Author: R.Krishna Prakash, rkrishprak@gmail.com

add the sender of the spam to the blacklist and the entire domain of the sender will be added to the blacklist. Currently received e-mails are checked, and if the sender is present on the blacklist, then e-mail are automatically considered as spam[4]. As with the whitelist, there are flaws with blacklists too.

Drawbacks of Existing System- The vital problem starts from the reality that spammers tend to forge header message in their spam[9]. The sender information is predominantly forged, that is may be end user's are added to a blacklist but more essentially the effect which the blacklist will have is diminished dramatically [6].

3. PROPOSED SYSTEM

The textual content of e-mail messages can be seen as a unique case of text categorization based on Spam filtering, with the categories being spam and non-spam[9]. Naive Bayes (NB) was recommended due to it robustness in the text-classification domain and due to potential of easily implemented in a cost-effective decision framework [6]. Proposed system use either personal mails or spam mails by a two-tier approach of using filters which is trained [10]. E-mails are categorized as authorized mails by the legitimate mail filter may pass, while the other remaining e-mails are processed using spam filter [4][10].

The spam email detector requires a training set which should have a pairs of e-mails and labels [2] specifying whether an e-mail is a spam e-mail or not. Spam e-mails and ordinary e-mails are training set classification[2][6]. A collection of samples which has pre-categorized label values as the training set and is used to identify the accuracy of the text classifier called a Testing set [2]. By comparing the label value assigned by the classifier with the pre-classified label value, compute the classification accuracy of the text classifier [2].

Advantages of Proposed System

- Robust: Robustness in the text-classification in Naive bayes classifier algorithm[6].
- Low False Positive and Low False Negative: False Positive and False Negative are limited due to calculation of spam based on probability[8].
- Text Attachment are also classified: The Training set compare the mail attached in mail box and then classified[1].Thus spam mail cannot be sent via attachment by intruder.

4. MATERIALS AND METHODS

OVERVIE- When the spam message received it will cause large consumption of network bandwidth and storage space which in-turn slow down email servers[8]. Hazardous content includes Trojan horses, worms, Viruses and other malicious codes caused by Spam software [4]. Hence, many security researchers are focused on. Numerous anti-spam technical solutions is been proposed and deployed to overcome this problem [8][1]. The most common and simplest way to reject or isolate spam messages at the receiving server using Front-end filtering[4]. However, prior anti-spam tools were constant; for example using a fixed set of keywords blacklist of noticed spammers or white list of good email determines the spam messages [8]. Though this list based method reduces risk, they neglects to scale and adapt spammers trick which can be easily achieved by changing the spelling of sender's address every time or bypassing the content in spam filters [8][1].

One fine difference is that a false negative is considered as serious error than false positive because, significant e-mail was determined as spam and rejected. Naive Bayes approach used is considered as most addressed machine learning techniques to fix the spam [1].

Proposed was using Spam Filter by Naive Bayes appoarch and it determine exact spam messages [1]. Probabilities are assigned to each attribute based on its number of occurrences in the training corpus and it is used to categorize a message by applying Bayes' theorem[3].

MODULES

- Designing User Interface Form
- Mail Client
- Mail Server
- Spam Mail Filtering Method

MODULE DESCRIPTION

*Designing User Interface Form- D*esign a GUI (Graphical user interface) part for user interaction with the application for e.g. user needs to register their details which include user name, password, and personal details. After successful registration, user need to login with user name and password. User has three type of process with this application that is Compose mail, Inbox, Spam.

Mail Client- Users run applications on client called Pc's or workstation. Clients seek resources, such as files, devices, and even processing power is based on servers. According to the process user should fill in the information as request to send to the server**.** Thus server provides mail access permission to user. If login is successful, client verifies their spam or mail box or creates mail to some other client with their username. Client can send the message or attachment to another client.

*Mail Server- S*erver stores the client request as records in the database, each time the client login will be verified by comparing the records in its database. Server can play a role to transferring the mail to other clients.

Spam Mail Filtering/Mail detector Method includes

- Preparation of Testing/classifying Set: All emails in the testing set are pre-classified and mixed together [1].
- Email Preprocessing: All the emails are preprocessed leaving the main bodies only [7].
- Generating word maps: A word map is a list of words that appear both mails. One word map contains words are both in the given email and the spam vocabulary table. Another word map contains the given email and the email vocabulary table. These two word maps can be different since some words in spam emails may not be in ordinary emails according to the training set [7].

5. SYSTEM DESIGN

INPUT-Input design is paramount phases of the system design. It is the process in which the inputs are received in the system are planned and designed, to get significant information from the user, vanishing the information that is not required. The aim of the input design is to ensure the level of accuracy and also ensure the understanding ability of users. Output will produce the error when system input is wrong[7]. *The goal to be considered during input design are:*

- Essence of input processing operation.
- Easy and complete of validation rules.
- Accuracy and coherence of input files
- Cautious design of input lead to good process lest errors will encounter

The input design of the system includes the following:

- Database Configuration
- Server IP Configuration
- New User Registration
- Login Form
- Compose Mail

OUTPUT- To communicate to the users the output design is been designed. Several outputs being designed are used to epresent the same format and procedure that the company and management used to. Direct source of information to the user is output of computer. Reliable, accurate and robust output design should improve the systems relationships with the user and help in decision making... A subtle point for the output design is the System Flow Diagram (SFD). Human factors reduce issues for design involves addressing internal controls to ensure readability [10].

- View Inbox and View Spam

DATABASE DESIGN

The goal of database design is to generate a set if relations that allows storing information easily[8]. The database is designed in relational model in which the data are organized into entities and relational between them. The following are the tables used in the system.

FIELD NAME	DATATYPE	DESCRIPTION
mailIndex	Int	Primary key to access
FName	Varchar	First name of the user
LName	Varchar	Last name of the user
userData	Varchar	User name of the user
Secure	Varchar	Password of the user

Table 4.1 New User Registrations

Table 4.1 is used to store the details of new user. It is used to identify the user when attempt to login.

FIELD NAME	DATATYPE	DESCRIPTION
fromAddress	Varchar	User name of the user who sends to another user
ToAddress	Varchar	User name of the user who is going to receive
AttachPath	Varchar	Path of the file in the attachment
Subjects	Varchar	Content in the subject
Message	Varchar	Content in the message

Table 4.2 Inbox Mails

Table 4.2 is used to store the mailing details of the user to user. The mails which are stored in this table are inbox mails.

SYSTEM FLOW DIAGRAM

System Flow Diagram is a graphical representation of the flow of data through an information system.

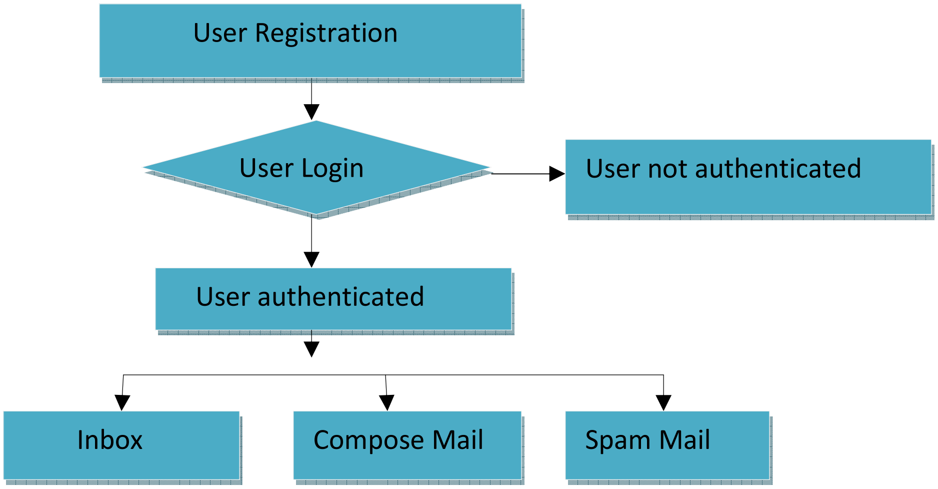


Figure 4.1 User Process

Figure 4.1 shows that the user need to register their details. The user will login using his/her login credentials. If the username and password matches, then the user is authenticated to view his/her spam mailbox, inbox and to compose new mail.

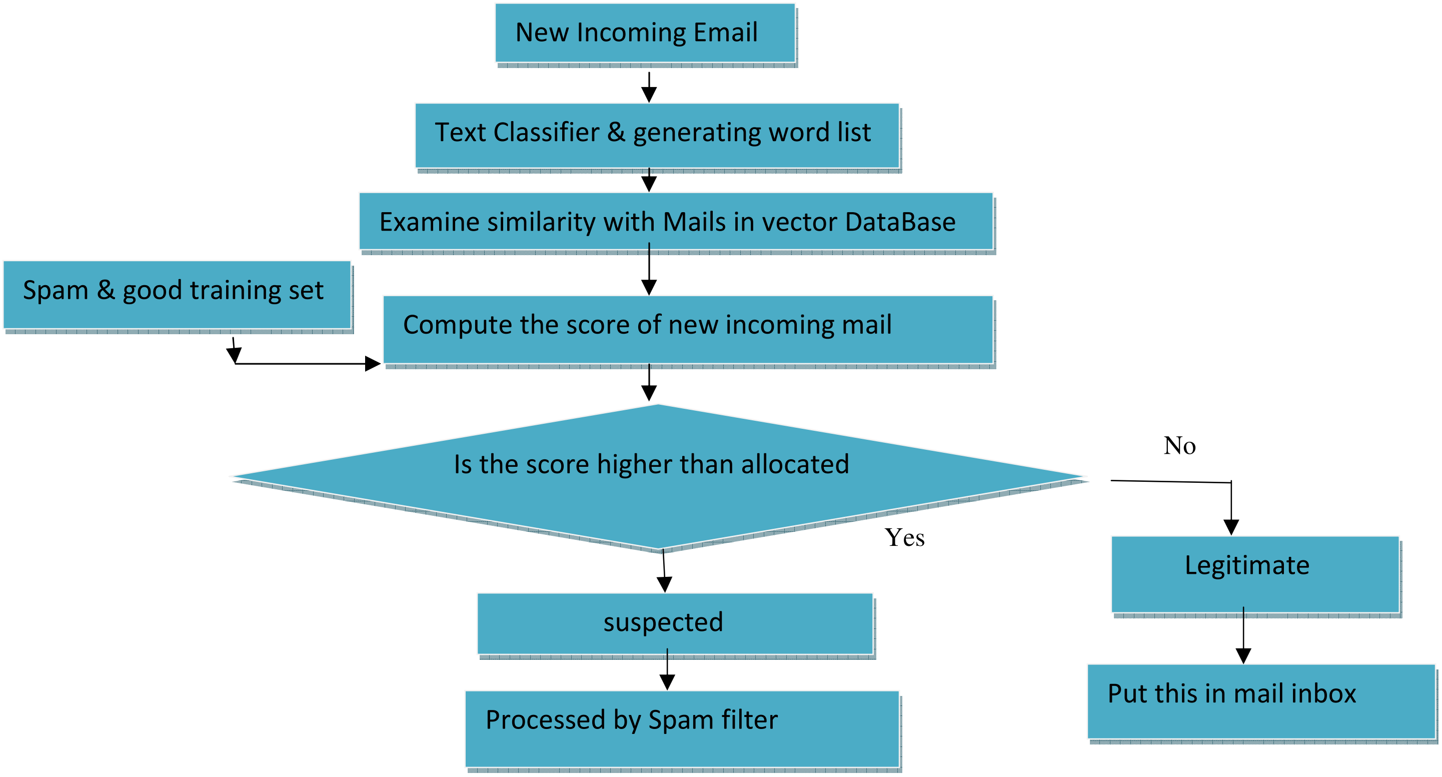
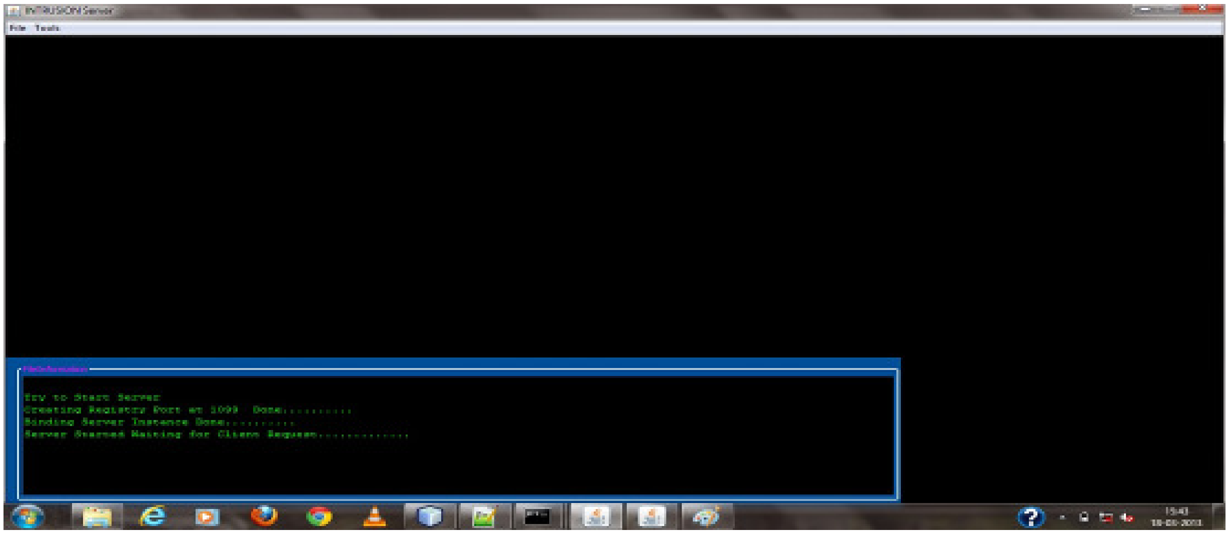

New Incoming Email
Text Classifier & generating word list
Examine similarity with Mails in vector DataBase
Spam & good training set
Compute the score of new incoming mail
No
Is the score higher than allocated
Yes
Legitimate
suspected
Put this in mail inbox
Processed by Spam filter

Figure 4.2 Naive Bayes Classifier Process

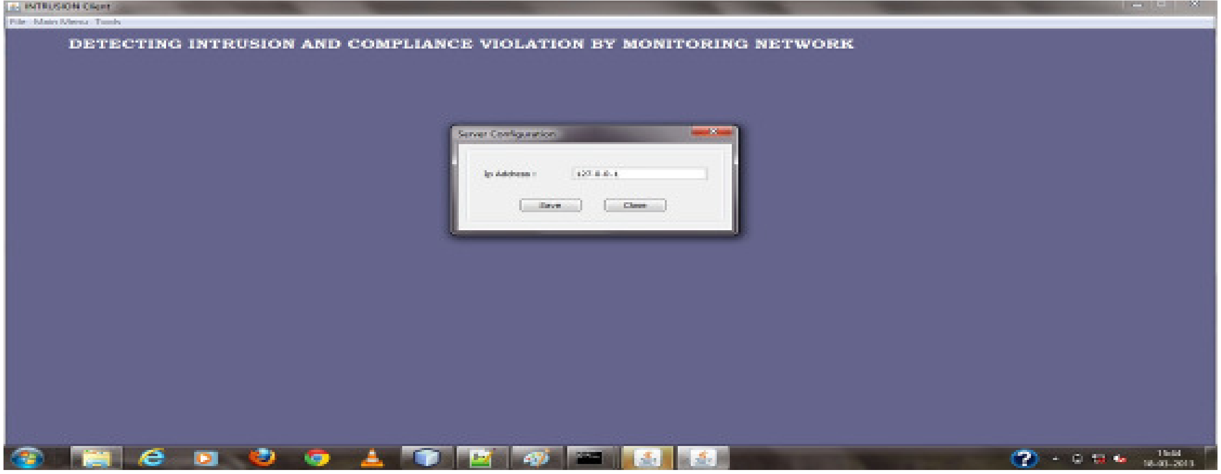
Figure 4.2 Naive Bayes Classifier Process describes about the flow of spam filtering process. Based on the probability the mail is filtered to inbox and spam.

6. RESULT AND DISCUSSIONS



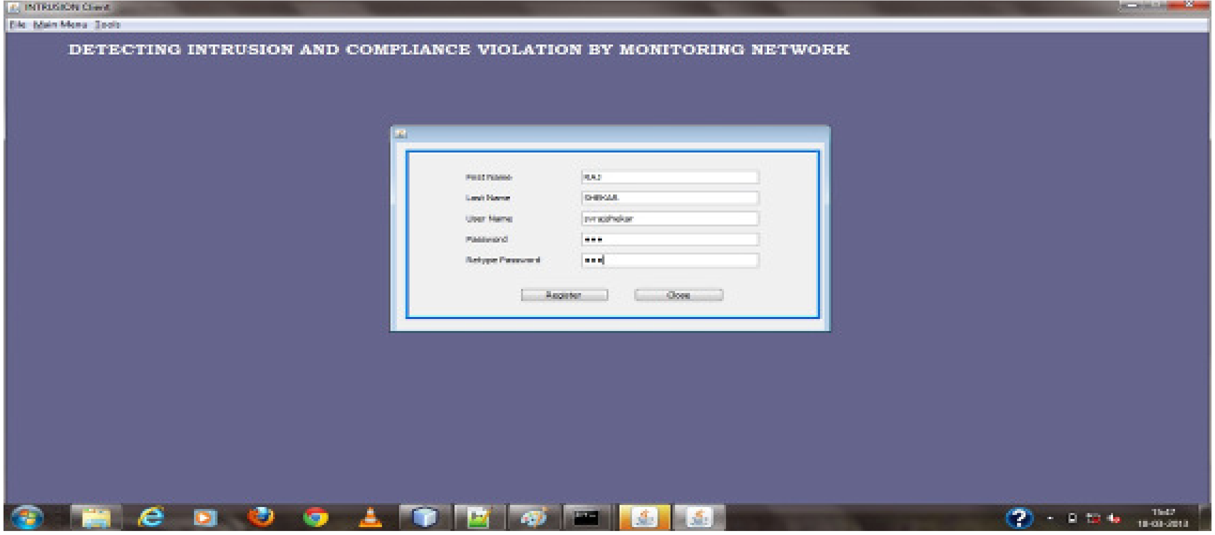
```
try to Start Server
Creating Registry Port at 1099  Done...........
Binding Server Instance Done...........
Server Started Waiting for Client Request..............
```

Figure.5.1: Database Configuartion:
The server is initialized first, the server waits for the client's request and responses to the request.


INTRUSION Client
File Main Menu Tools
DETECTING INTRUSION AND COMPLIANCE VIOLATION BY MONITORING NETWORK
Server Configuration
Save
Close

Ip Address : 127.0.0.1

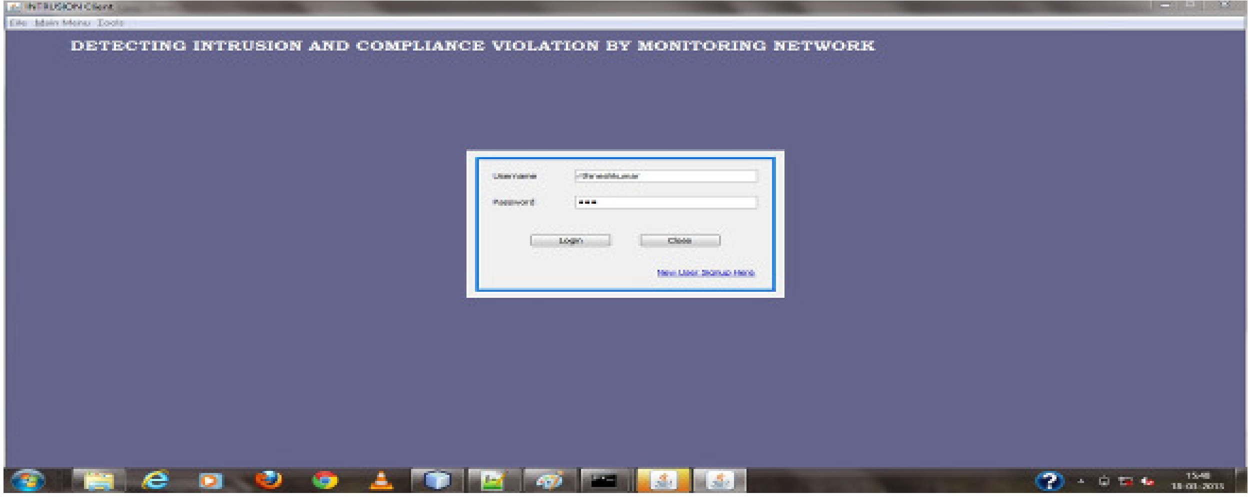
Figure 5.2: Server IP Configuration
The IP address of respective users will be entered here to ensure the correct user


INTRUSION Client
File Main Menu Tools
DETECTING INTRUSION AND COMPLIANCE VIOLATION BY MONITORING NETWORK
Register
Close

First Name RAJ
Last Name SHEKAR
User Name rvrajshekar
Password ***
Retype Password ***

Figure 5.3: New User Registration

The following interface allows new users to create account. Users have to mention their firstname, lastname, username, password

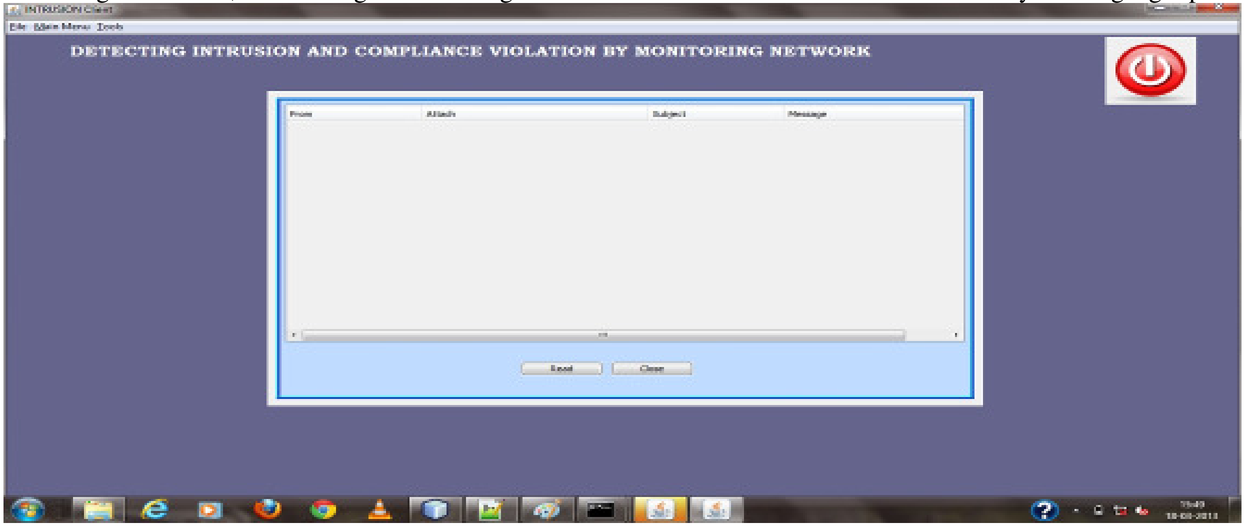

INTRUSION Client
File Main Menu Tools
DETECTING INTRUSION AND COMPLIANCE VIOLATION BY MONITORING NETWORK
Login
Close
New User Signup Here

Username rdhineshkumar

Password ***

Figure 5.4: Login Form

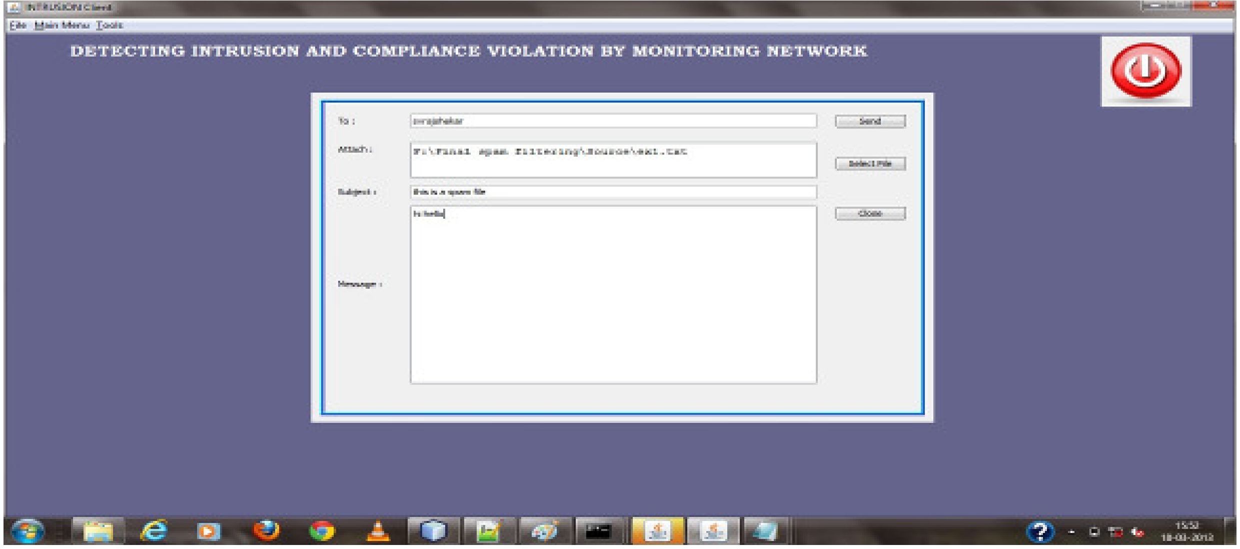
The login interface, the existing users can login here and new users can create new account by clicking signup label.


INTRUSION Client
File Main Menu Tools
DETECTING INTRUSION AND COMPLIANCE VIOLATION BY MONITORING NETWORK
Read
Close

From	Attach	Subject	Message

Figure 5.5: Intrusion detection Form

Inbox view of a user. The user can view their good e-mails here. Intrusion detection will be carried out as well.


INTRUSION Client
File Main Menu Tools
DETECTING INTRUSION AND COMPLIANCE VIOLATION BY MONITORING NETWORK
Send
Select File
Close

To : rrajshekar

Attach : F:\Final spam filtering\Source\ext.txt

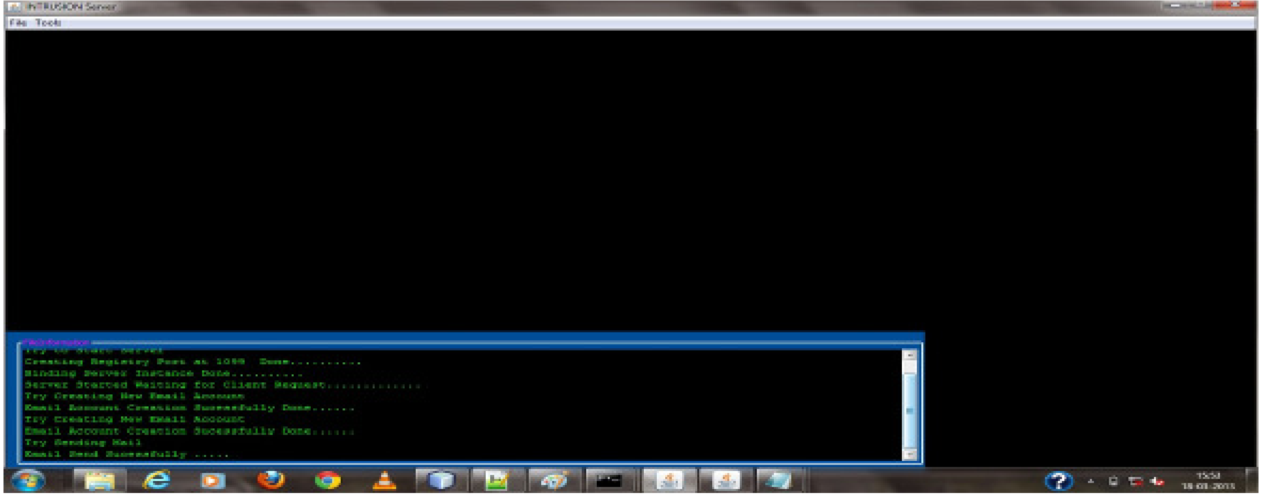
Subject : this is a spam file

Message : hi hello

Figure 5.6: Compose Mail

View of the compose mail interface. Here we can compose either spam or good e-mails.

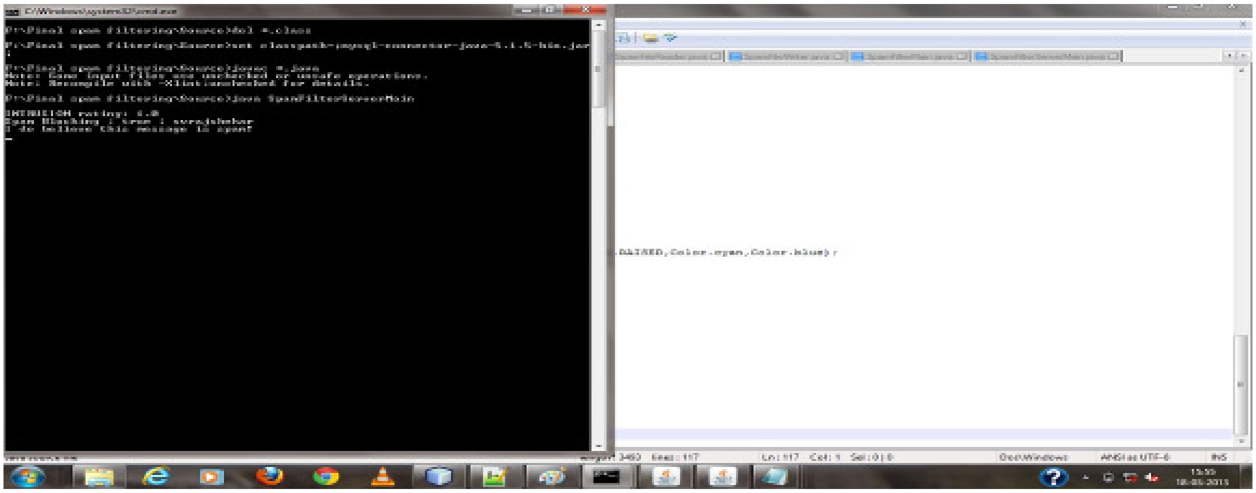
In the following snapshot, the server is responding to the client's request



```
Try to Start Server
Creating Registry Port at 1099  Done..........
Binding Server Instance Done..........
Server Started Waiting for Client Request..............
Try Creating New Email Account
Email Account Creation Sucessfully Done......
Try Creating New Email Account
Email Account Creation Sucessfully Done......
Try Sending Mail
Email Send Sucessfully .....
```

Figure 5.7: Server Response

The sent e-mail has a spam, the Naive Bayes Classifier classifies it as spam. The following snapshot shows spam rate and sender's name.

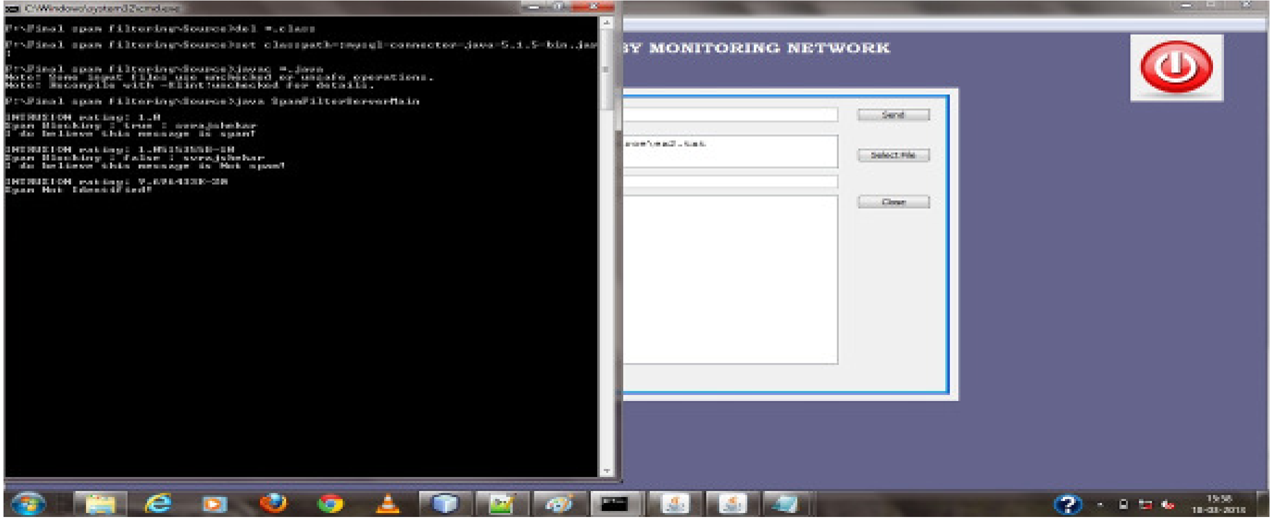


```
F:\Final spam filtering\Source>del *.class

F:\Final spam filtering\Source>set classpath=mysql-connector-java-5.1.5-bin.jar
;

F:\Final spam filtering\Source>javac *.java
Note: Some input files use unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.

F:\Final spam filtering\Source>java SpamFilterServerMain

INTRUSION rating: 1.0
Spam Blocking : true : swrajshekar
I do believe this message is spam!
_
```

```
.RAISED,Color.cyan,Color.blue);
```

Figure 5.8:Spam Mail

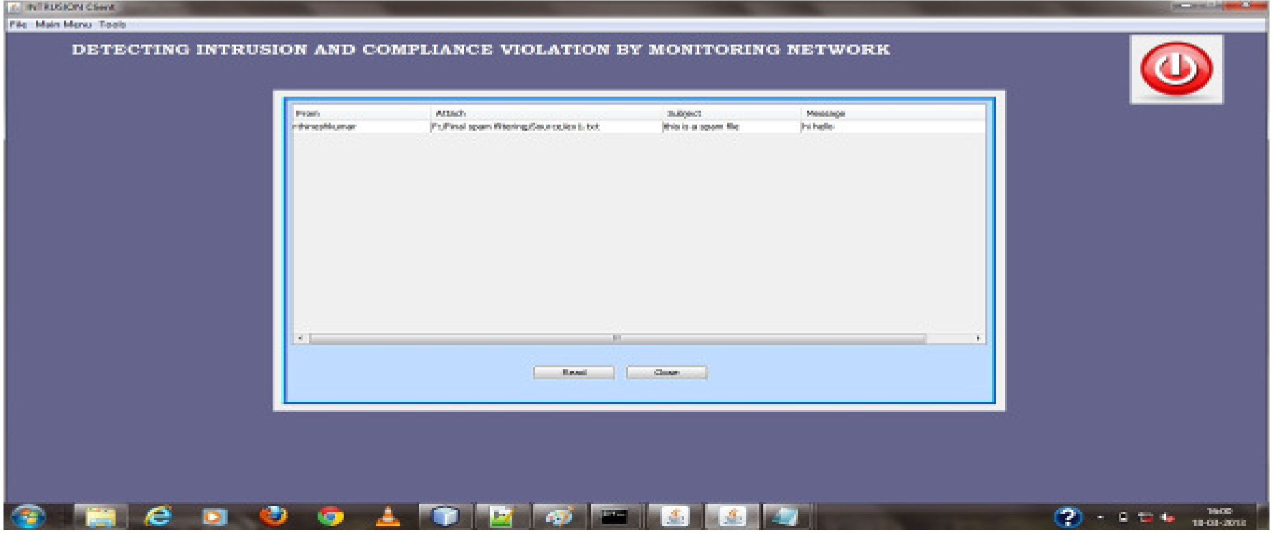
Following Snapshot shows that the spam rating of particular e-mail is low and hence spam not identified.



```
F:\Final spam filtering\Source>del *.class

F:\Final spam filtering\Source>set classpath=mysql-connector-java-5.1.5-bin.jar
;

F:\Final spam filtering\Source>javac *.java
Note: Some input files use unchecked or unsafe operations.
Note: Recompile with -Xlint:unchecked for details.

F:\Final spam filtering\Source>java SpamFilterServerMain

INTRUSION rating: 1.0
Spam Blocking : true : swrajshekar
I do believe this message is spam!

INTRUSION rating: 1.0515355E-18
Spam Blocking : false : swrajshekar
I do believe this message is Not spam!

INTRUSION rating: 9.6914338E-28
Spam Not Identified!
```

Figure 5.9: Calculation of Spam

The following snapshot is the spam mailbox view. The mails which are considered as spam are moved to spam mailbox.


INTRUSION Client
File Main Menu Tools
DETECTING INTRUSION AND COMPLIANCE VIOLATION BY MONITORING NETWORK
Read
Close

From	Attach	Subject	Message
nthineshkumar	F:\Final spam filtering\Source\ss1.txt	this is a spam file	hi hello

Figure 5.10.Spam mail moves to Spam box

7. CONCLUSION

The main objective of this system is to apply Naive Bayes algorithm on email classification. Bayes Theorem and Naive Bayes algorithm for text classification are efficiently used. Explored methods are also used for text preprocessing and probability computation in text classification. However, there remains scope for improving the performance of the spam e-mail detector. The performance is empirically evaluated using a public corpus and compared its effectiveness with a Naive Bayes classifier. It was shown that the classifier can achieve very good results provided that choose enough training samples of legitimate and spam e-mails.

The spam e-mail detector uses a static model with probabilities of words unchanged during the classification step. Such a model is restricted due to the flaws of prior knowledge may cause noticeable inaccuracy in spam email detection. Statistically, the bigger the training set is and the more randomly that emails are chosen, the better model can be obtained. However, the size of the training set is always limited and the randomly picked emails only reflect the trend of spam e-mail in a given period. Thereby, a more comprehensive and robust model can be constructed by tuning probabilities of words in the vocabulary table dynamically during the classification step. The advantage of such a dynamic tuning mechanism is that it enables the model to reflect the latest trend of spam e-mail, which eliminates the cost of manually adjust the model or even rebuild the model.

REFERENCES

[1]. Androutsopoulos, J. Koutsias, V. Chandrinos, and D. Dpyropoulos, ‘An experimental comparison of Naive Naive Bayes and keyword-based anti-spam filtering with personal e-mail messages’ In 23rd Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval, ISBN:1-58113-226-3, page no.160-167 , 2000.

[2]. Blum, T. Mitchell, ‘Combining labeled and unlabeled data with co-training’, in Proc. Workshop on Computational Learning Theory,ISBN:1-58113-057-0, page no.92-100, 1998.

[3]. Daniel Grossman, Pedro Domingos University of Washington, Seattle, WA ‘Learning Naive Bayes network classifiers by maximizing conditional likelihood’ 21st international conference on Machine learning table of contents Banff, Alberta, Canada, Learning (IDEAL04), UK ISBN: 1-58113-838-5, page no.361-368, 2004.

[4]. H. Drucker, D. Wu, and V.N. Vapnik, ‘Support vector machines for spam categorization ’IEEE Transactions on Neural Networks, vol. 10, no. 5, page no. 1048-1054 , 1999.

[5]. Jonathan Palmer, ‘Naive Bayes Classification for Intrusion Detection using Live Packet Capture’, data mining in bioinformatics, 2011.

[6]. J. Provost ‘Naive-Bayes vs. rule-learning in classification of e-mail’ The University of Texas at Austin, Department of Computer Sciences Rep, AI-TR .99-284, 1999.

[7]. M Rogati, Y Yang, ‘High-Performing Feature Selection for Text Classification', CSD, Carnegie Mellon University, CIKM’02, ISBN: 1-58113-492-4, page no.659-661, 2002.

[8]. G. Sakkis, I. Androutsopoulos, G. Paliouras, ‘A memory-based approach to anti-spam filtering,’ Information Retrieval, vol. 6, no.1, page no. 49-73, 2003.

[9]. F. Sebastiani, ‘Machine learning in automated text categorization’ ACM Computing Surveys I vol. 34, no.1, page no.1-47, 2002.

[10]. X.-L. Wang, I. Cloete, ‘Learning to classify e-mail: A survey’ In Proc. of the 4th Int. Conference. On Machine Learning and Cybernetics, Guangzhou.vol. 9, ISBN: 0-7803-9091-1, page no.5716-5719, 2005.