Project Final Report

(Submission Date: 18 Feb 2019)

E-MAIL SPAM DETECTOR

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

Submitted By:
ARINDAM(17BCS1780)
SURAJ(17BCS1769)
RAHUL(17BCS1811)

Name of the Mentor:
Manpreet Kaur

Roll No:
17BCS1780
17BCS1769
17BCS1811

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Chandigarh University, Gharuan

PROJECT GOAL

The goal of the project is to design an E-mail spam detector which detects the spam in e-mails and it sends to the trash box in spam section, It determines the spam e-mail itself as its designed in such a way which we call self automation(.i.e. which detects the spam e-mails itself). As soon as we receive the spam e-mail it starts functioning and informs us about the spam email server automatically. It is coded in python programming language.

ABSTRACT

Identifying and fixing the affected machines is the key step to resolve any security threats in a network. Because, it becomes a route to launch several attacks such as Denial of service attacks, spamming, stealing user identities and spreading malware etc. Spamming is one of the major threats where attackers perform single attack and make multiple machines in a network as compromised machines. Even though few existing methods like spam signatures and spam behavior analysis resolved the problems to certain extent, it is still not applicable in large networks. Moreover, these methods lack online spam detection mechanism. Existing systems and its drawbacks are also discussed in network administrator. A definite algorithm in this report is used to differentiate between spam and non-spam. The performance of this tool is based on the parameters like number of spam messages, percentage of spam detected and its efficiency to overcome the limitations of the existing systems.

Table Of Contents

LIST OF FIGURES

INTRODUCTION

Due to the wide popularity of the internet and its communication with no cost, it was recognized as the premium tool for advertising and marketing. With respect to economic constraints, most number of people started sending emails to thousands of people across the world. This made internet, a commercial network with the association of electronic mail as one of the quick resources of communication. The major problem in today's internet world is sending bulk or unsolicited emails to numerous users. This adds an additional advantage of launching other attacks and wasting of resources [1]. E-mail spam comes under the electronic spam which sends bulk of unnecessary or junk mail of duplicate emails to the recipients.

1.1 Types of spam:

1.1.1 EmailSpam:

Email spam is the most familiar spam that most of the users come across every day. Email spam follows three properties i.e., anonymity, mass mailing and unsolicited emails. Anonymity is the property of hiding the uniqueness and whereabouts of the email sender. Mass mailing is defined as the sending of bulk identical emails to the large number of groups and unsolicited emails are the emails transferring to the recipients who do not request. Typically, an email sent to large number of groups without any requestbyhiding their identity is referred as email spam. [1][16]

CommentSpam:

This is the most common spam that many users come across in various blogs. Spammers use the posts in the blog to redirect to spam websites. The ranking of such blogs gets increased gradually in the search engines. It is basically used to promote the searching services like Wikipedia, blogs, guest books etc. There are number of tools in the market to get rid of comment spam.

1.1.2 Instant messengerspam:

It is not as widely spread as other types of spam. Yahoo messenger, My space, Windows live messenger etc. are the end spots for the spammers. The spammers gather the data of different users and send unsolicited messages within a link that triggers viruses, spam etc. The best way to get rid of this type of spam is to ignore the messages from the strangers. There is also a possibility to get the links from the existing friends list. A critical measure like verifying the size of the URL will be able to trim the chances of being the victim to this spam.

1.1.3 Junkfax:

Junk faxes are not as prevalent as before. It reduced periodically with the existence of internet technology. However, there are also some risk factors occurring in few corners because of this telemarketing technology. This is similar to junk email where the advertisements and messages are passed to numerous users via fax machines. The adversaries use broadcast fax as a medium to pass on the junk fax to various users. Fortunately, there are surplus tools to overcome junk fax.Unsolicited textmessages:

This is kind of similar to instant messenger spam but here the messages are passed via mobiles. SMS is the service through which the messages are transferred from one user to other user. The easiest way is to maintain the contact with the known friends instead of strangers. It is relatively easy to find the source where the message is coming from with the instant messenger spam. It is critically important no to click on the links that are passed via mobile by the spammers.

1.1.4 Social networkingspam:

Social networking sites play an important role in today's world. With the advent of such sites, spammers also started flooding using new techniques to make the social networking sites such as face book, twitter, linked in etc. as part of the spamming activities. As of now it is targeting only the wall posts, messages but these techniques evolve certainly over a period of time. Spammers use notes or messages through various groups or pass the messages with embedded links, which may lead to pornographic or other sites and target spam [4]. Even though these sites have an option to report spam or abuse activities, the spammers frequently change their address or account to hide their identities.**Problemswithspam**:

1.1.5 Viruses:

Viruses are the most dangerous threats across the network. There are many techniques and methodologies developed to decrease the nefarious activities caused by different types of viruses. With the increase in the internet technology, wide variety of viruses produced to attack the machines. Spam is one of the sources to launch such types of viruses. The widely spread viruses are the ones which disconnect the hosts and get diffused into the network. Spam viruses in modern technology are more dangerous as it controls the machine itself and then annihilates them. Viruses are not visible and get launched when a particular command is triggered. There are so many techniques used by spammers in order to allow users to click or use the links to launch thousands of spam viruses across the network. Due to increase in the intensity of spam, it captures the user's email address and passes numerous messages to the customer list, through which it disturbs the customer trust and destroys the system. [4]

1.1.6 Serverproblems:

Most of the time servers are being targeted by the spammers. Due to increase in the intensity and volume of the spam, the company or any system has to use huge resources to maintain the server. In order to distill and disseminate the data that is transferring in the network more energy costs and resources are to be divided among the departments. Due to this frequency of spam, the performance also gets affected. So, the servers must maintain a low and necessary data. Otherwise, it can create major problems on the server to maintain and causes heavy load disrupting the entire network.Hacking and Phishing:

As the computers in the modern technology are becoming more and more secure, the spammers face more difficulty to capture the confidential details. So, they tend to use various methods to break through the security of different IT departments. Spammers make use of hacking methods like entering into the trusted employee system without the user's awareness. Then, spammers perform different activities and keep a record of the confidential data or hold vital information either for the cost or for self-happiness. Another way is to trap the employees of the companies to enter the passwords or any valuable information into the spammer's website, so that it keeps track of the password to reveal important credentials [7]. Though there are many firewalls and spam filters, spammers are also improving their technical skills to intrude on organizations.

1.1.7 Productivity threats:

It is known fact that most number of employees in any organization spends approximately an hour of time to sort out and delete the spam from a cluster of good emails. This leads to heavy wastage of resources like labor cost, time and space in any system. An important email among the cluster of non-spam emails seems to be an unimportant one. This causes problems like loss of e-mail, deleting email and can also disturb the valued customer trust and internal correspondence.Blank spam emails and forwarding spamemails:

Spammers also use the technique of sending a blank email to the recipients. The purpose of sending this type of email is to recognize whether the recipient possesses a valid email ID or not. If it is an invalid email ID, then it bounces back stating with a non- deliverable notice. This helps the spammer to identify whether it is a valid email ID or not. Sometimes blank emails also attach few files that initiate Trojan virus if it is opened in the system. Forwarded emails are again one more cause to initiate spam emails. This makes users forcibly to forward the users in their friends list. As a result it forms a chain and delivers spam emails and becomes uncontrollable. This eats away lots of time and space and costs a lot to filter spam in the system. [19]

1.2 Types of email spamfilters:

Spam filter is a piece of software that is used to filter the spam emails based on the content and rules adhered by its corresponding software. Every single spam filter has its own set of rules through which the spam is filtered from spreading across the network. It involves the content of the spam, address of the users and where it is redirecting to etc. Based on these parameters it judges, whether an email is a spam or not. There are multiple spam filters divided based on their rules[21].

1.2.1 Challenge-Response spamfilter:

This spam filter is a basic filter mechanism that is used to control the spam in the emails. This does not allow any strangers or any pre-approved persons to send an email to the user. In return to the email sent by these pre-approved persons, it asks to validate them in order to pass on the email. The logic behind this strategy is that approvedusers do not have time to validate their own email ID from thousands of emails that it might have sent. However, there are numerous problems with respect to this system.

- There are high chances that the spammer uses its fake address in order to validate email address. So therefore, even challenge responses are sent, it can be able to validate itself based on the modified address. So there will be exchange of messages and can cause spreading of spam due to a spammer validating email address.
- Another major problem is during the online selling or buying products. Whenever a product is purchased, a receipt or the details of that particular product may be sent to personal email through an automated email box. Here the problem arises, causing an automated mail box to verify its own email ID. As this mail box cannot reply and validate itself, this email is discarded causing disturbance in the online purchase.
- The spam filter itself can create a problem to the other user. This is a case when an email is sent back to an actual person mentioning to authenticate the valid user. The spam filter on the other side may also stop the email that is sent by spam filter from entering into the inbox as it is not sent by the intended recipient.
- This filter requires the user to crack the code whenever an image is to be transferred via email. This perhaps becomes a problem for the users with

disabilities. Even though entering the code is a good cause to determine whether it's the actual person or any part of the software, it may be troublesome for the users to break the code every single time just to send an email. So there are possibilities of not sending an email. In the business point of view, this becomes a major problem as the email is sent to any other competitor avoiding the person containing spam filter as it is difficult to send an email.

This filter filters the spam to certain extent, but not completely. Even though it could help in filtering few emails, it has its own drawbacks, which can cause disturbance in the network.

1.2.2 Rule based scan filtering system:

These are referred as the original spam filters. It works on the method of detecting pre-determined words or phrases that most of the spammers use. It identifies those key words and block's emails from passing on from one user to other users. The rules to detect words or phrases are to be improved daily. Because the spammers are so intelligent that it keeps track of words that are blocks and uses its synonyms for a successful transmission. Nevertheless, strict based rules also become another problem, as it blocks even a legitimate email. There is every possibility that both spam and non-spam email get blocked due to the demanding rules passed on by the rule based to scan filtering system. Suppose there is a word spam involved in the message, it gets blocked by the rules based on the filtering system. If the rules are really weak, then spammer' tries to modify the message from "spam" to "sp@m" and passes on the message successfully and spreads the spam across the network. So, there needs to maintain a different strategy for the rule based scans. Even so, one has to accept the fact that the effects of spam can be reduced,

1.2.3 Global black lists spamfilter:

Global black lists contain the list of the notable spammers. Whenever an email is sent, the internet keeps track of the email sender details, i.e., from where the email has come from. Its IP address also gets recorded. So this spam filter compares against the black list containing the notable spammers. If there is a match, it discards the message before reaching it to the recipient. This makes users to escape from the notable spammers. The black list also gets updated so that it can reduce the well-known spammers to perform the spam activity again. This does not waste its space and time by verifying repeatedly with the spammers but perform only one search with just one database thus saving lot of time and space.[7]

There are certain problems with this filtering system. The black list is decided by the users familiar with detecting spam. So before a spammer getting into this list, there happens to transfer thousands of emails to the users. Sometimes, the legitimate person may get into this black list thus getting boycotted completely without any illegitimate activity. And also, complete internet service providers are blocked because of few users involve in spamming activities. This results in disrupting the entire network.

1.2.4 Bayesian Analysis:

All the methods that were discussed above are based on some predictive methods and content in the message during the exchange of information. However, this filtering system is based on the mathematical formulae through which the email can be determined if it is a spam email or non-spam email. As the black lists take a lot of time to get it self updated,the spammer sork. So, in this filtering system it just uses a sample of data and determines if the email is a spam or not in short duration.

This is also associated with set of problems. This formula is developed long ago. So, there is every possibility that the spammers get updated and escapes with the sample of data that is used in the formulae. The key point is to know how properly the formula is developed and how well it is reliable. Implementing the same formulae after so many years is also a not good idea.

1.2.5 Permission based spamfilter:

In this type of filter, user will be given all the permissions whether to send emails to inbox or trash the emails. All the permissions on the emails can be customized by the user. The privileges are created by the user based on the content of the data, header information etc., to allow the transmission of data from one user to another user.

1.3.6. Existing models

Many models have come into existence in order to reduce the amount of email spam transfer across the network. Despite the awareness of spam emails, productive efforts have not been developed for the network administrator to monitor the status of clients in the network. Each model has its own advantages and disadvantages. Following are few existing systems and their drawbacks.

2.1 Spam signature generation based frame work:

The aim of this model is to analyze the behavior of spam in the network through its characteristics and properties. This behavioral analysis identifies the spam loop holes in the network thus helping the future strategies for the prevention of spam spread across the network. This mechanism is developed based on the framework AutoRE which identifies the spam attackers based on the signatures from the exchange of messages. This framework is based on the signature based worm and virus detection systems.

- According to this framework, a bunch of spam emails are sent across the network to study the characteristics of spam behavior. It mainly targets on the emails that have embedded URLs as it is one of the direct sources of targeting to spammer's location and phishing websites.
- Separating the legitimate URL with illegitimate ones is a critical task in this framework. Most of the spam emails contain several URLs out of which some are general and non-general. Spammers use the techniques to avoid detection between the legitimate and illegitimate emails. AutoRE uses the bursty property bias electing spam URLs in order to encounter the challenge of separating goodandbad URLs. This framework does not use any black lists or white lists as it generates regular expression through which the spam URL is compared and filtered.
- Figure 1 represents URL preprocessor, group selector and RegEx generator are the three modules used in this framework. The task of identifying URLs and other relative details from the emails of the end user are clustered based on the web domains. Based on the degree of burstines, group selector selects the URL clusters that were performed by the URL preprocessor and transfers to RegEx generator. This module identifies the signatures by analyzing cluster behavior and generates for all the clusters. The signatures that match with the URLs are separated from the other URL groups to evade further progress. This is repeated until all the clusters are fulfilled.

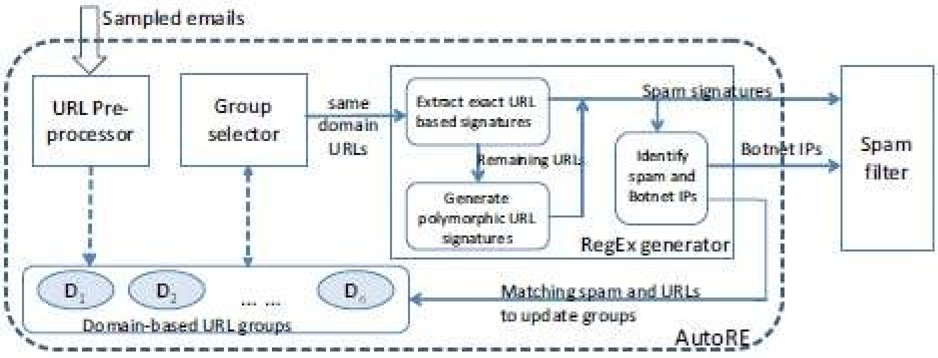


Figure1: AutoRE modules and processing flow chart [8]

- Regular expression signatures are more robust relative to the fixed string based signatures and able to identify more spam information. These signatures can highly bring down the false-positive rates in identifying the URLs. Moreover, this framework also uses these signatures to the large number of group emails, which are referred as spam campaigns. This framework used some email patterns from hot mail and monitored for three months in order to study the spam characteristics [8]. The success ratio of this framework is really high based on the studies it carried out on the email traces.

Disadvantages:

Even though this frame work is accessible for the URL groups embedded in email, it lacks in action taking for IT security groups. The degree of burstiness in the real time is one more major drawback that is faced by this frame work. It does not provide any clarity about how well it is performed in a real time for the spam campaigns. Spammers can easily develop techniques to meet the preventive measures of AutoRE framework like making legitimate domains fall into the list of illegitimate emails. The results that are obtained do not provide any aggregate view of the large groups of emails. Moreover, it does not let the network administrator to have an online monitoring system across the network.

2.2 Characterizing botnets from Email SpamRecords:

This framework presents techniques, which use traces of spam email to map botnet in groups. This is done by viewing for several bots involving in the same spam campaign. This has been used against a sample of spam email from Hotmail web mail system and has successfully detected multiple botnets. This technique uses a large set of spam emails as input, which are destined at Hotmail in a regular period. Group of botnets that involve in launching spam emails with respect to its statistics like dimensions and events are observed in the output. Three major steps are involved in identifying them. With "clustering email messages," spam email messages with identical content are transferred from the entity which is controllable [12]. Firstly, the spam campaigns are to be identified. Most of the spam content in the campaigns appears to be similar, but in order to evade the differences; identical properties are to be considered. For efficiency, shingling concept is used [2]. A unique property like finger print is considered here and counts the number for every single message. Based on the number of finger prints that share among the spam campaigns identical nature is computed.

The second step involves "Assessment of IP dynamics" where the results are dependent on the dynamic IP addresses of the hosts. But, for this approach where the results should affect the dynamic IP addresses, different segment of IP address space are to be used for the computation of IP based parameters. For every C-subnet, two parameters are to be used through which it calculates whether two spam messages occur from the same machine at distinct times. The two parameters include average IP address reassignment time and IP reassignment range.

The final step is "Spam campaigns to botnets", where it is assumed that group of spam email messages is possible to join from the identical spam network. Depending on the first two steps, it combines every spam campaigns into a set of spam campaign that is originated from the same network of bots. Based on the statistics of the IP dynamics, for each corresponding message in a spam campaign, even the sending host involves in the consecutive spam campaigns. Two spam campaigns are also combined if huge numbers of participants are involved.

Disadvantages:

This framework deals better with the estimation of sizes and activities based on the results generated through common characteristics. Still, it lacks in validating these results against some other online techniques like IRC. Comparison always provides a better view and shows the correctness and success ratio of the framework. Spam email messages that comprise of images can be dealt using group of shingles. Most importantly, as in the previous framework it lacks to provide online detection and monitoring of the networks.

2.3 Botsniffer:

Botsniffer is defined as "a prototype system which is used to capture the spatial-temporal-correlation in the network traffic and utilize statistical algorithms to detect botnets with theoretical bounds on the false-positive and false-negative rates."

Botnets here are the network of zombies and are recognized as a very serious threat these days. Command & Control (C&C) channel is one of the main characteristics of botnets when compared to other malware. They use certain protocols like, for example: IRC, HTTP.

In order to identify the botnet C&C in local area network, an approach is proposed that uses network based anomaly. This is mainly based on the observation of spatial-temporal correlation and similarly which bots within the same botnet and demonstrates without any knowledge of signature and C&C server address in prior [15]. The Architecture of a botsniffer is illustrated in Figure 2. The botsniffer has two main components

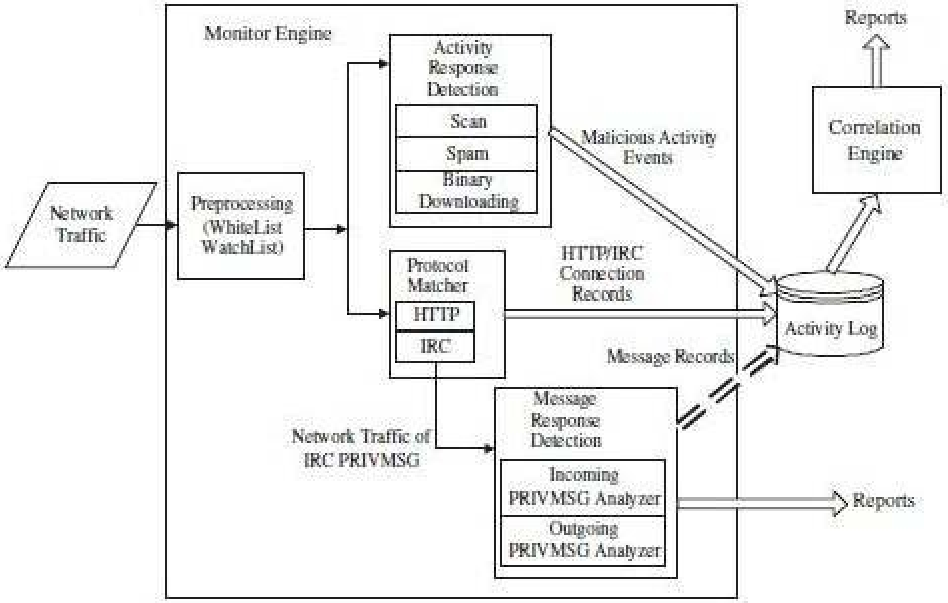

Monitor Engine
Network Traffic
Preprocessing (WhiteList WatchList)
Activity Response Detection
Scan
Spam
Binary Downloading
Malicious Activity Events
Protocol Matcher
HTTP
IRC
HTTP/IRC Connection Records
Message Records
Activity Log
Correlation Engine
Reports
Network Traffic of IRC PRIVMSG
Message Response Detection
Incoming PRIVMSG Analyzer
Outgoing PRIVMSG Analyzer
Reports

Figure 2: Bot sniffer architecture [6]

2.3.1 Monitor Engine:

2.3.1.1 Preprocessing:

Botsniffer filters the unnecessary traffic to lessen the heavy traffic when the network traffic steps into the botsniffer monitor engine. It filters out the protocols that are likely not used for C&C communications, such as ICMP and UDP. It also uses a white list to filter out traffic to general servers. This normal server's list is generated after gradual analysis. Watch list here is generally not necessary, but it helps in studying the network characteristics of the local clients.

& C - like protocol Matcher:

A client record is maintained using C & C like protocols for the benefit of correlation. IRC and HTTP are the two paradigms on which this matcher is very much dependent. Port Independent protocol matchers were generated to find all suspicious IRC and HTTP-related network traffic [3]. The IRC as well as HTTP protocols are relatively easy to identify with connection registrations.

2.3.1.2 Activity/Message Response Detection:

Botsniffer monitors the client's network behaviors that are involved in IRC or HTTP for signs of bot responses like message response (e.g., IRC PRIVMSG) and activity response behavior (e.g., Scanning, spamming). [6]

2.3.2 Correlation Engine:

The Clients are made into groups according to their destination address and port pair in this phase. Clients fall into the same group if they are connected to the same server. After this phase, botsniffer analyses the group behavior of spatial-temporal correlation and similar characteristics [5]. Afterwards it issues botnet notifications if it detects any doubted C&C. Botsniffer uses the response-crowd-density-check algorithm for group activity analysis and response - crowd-homogeneity-check algorithm for group message response analysis. If there are any notifications from these two algorithms, it subsequently generates a botnetreport.

2.3.2.1 Response-Crowd Density CheckAlgorithm:

This Algorithm verifies for any dense crowd response to the power of 2 for each time window. As a sub group, a lookup is done for any message or event response. If any number of clients with the activity behavior in a group is larger than a threshold (e.g., 50%), then it can be referred as the response coming from the dense crowd.

If several response crowds are considered, the situation of two cases arises. One is either an element of botnet and the other is not an element of botnet. Furthermore, to decide the number of response crowd required, a SPRT (Sequential Probability Ration Testing) algorithm is used to reach a decision within minimal rounds and with a bounded false positive-rate and false-negative rate. [11]

Disadvantages:

- It requires observing multiple rounds of response crowds. If there are only a few response crowds the accuracy of the algorithm maysuffer.
- Sometimes not all bots respond within the similar time window when there is a relatively loose C &C.
- It does not provide aggregate global characteristics of spam botnets involved in spamming.
- It lacks in providing an automated view of compromised machines in a network in an online manner.

2.4 Botminer

Botnet is compromised machines under the influence of malware code [10]. These compromised machines works under the influence of boot master and utilizes all the resources to counter denial of service (DOS), spam attacks, phishing, and identity theft. Command and control channel (C&C) is used by the boot master to issue the commands to the bots and to coordinate between different computers. Command and control (C&C) channel use the Internet Relay Chat (IRC) structures, in which the commands issued to the bots are independent of each other. IRC helps in avoiding collision between the bot commands. As command and control channels belong to centralized networks; they suffer single point of failure. To avoid this drawback, it uses peer to peernetworks.

Botnet architecture consists of five main components: A-plane monitor, C-plane monitor, A-plane clustering, C-plane clustering and cross-plane correlator. A-plane monitor and c-plane monitor are used to monitor the network traffic, and they run in parallel.

- A-plane monitor is used to monitor, who is doing what, A-plane monitor is capable of detecting malicious activities by scanning the outbound network. It detects the spamming, binary download etc. A-plane alone cannot detect all the malign activities in bot networks; it also generates false positives.
- C-plane records the information flowing in the network and determines who is talking to whom. The information recorded contains IPaddress, source, and destination in both the directions. C-plane networks are capable of recording data in high-speed networks.
- A-plane clustering is divided into two types of clustering. It contains the client list with illegitimate activities. These activities include spam activity, scan activity, binary downloading, exploit activity. The clients at least perform one activity in a day. To detect a spam activity both the clients should be in SMTP destination servers. C-plane clustering reads the logs generatedbyc-plane monitor. It identifies the clusters and machines and shares the communication. It is done in two ways; in basic filtering it directs the clients from internal hosts to external

hosts, and it ignores the communications between them. In white filtering it filters the destination points which are usually known as legitimate servers[10].

Disadvantages:

- Botminer may not be considered as the right choice to detect all the attacks accurately.
- It is easy to manipulate the communication patterns between botnetmembers.
- If the bot master tries to generate the new commands in any unspecified manner, it tends to promote previouspatterns.
- If botmaster tries to deploy the activities to stop the attacks, then it becomes difficult because communication channels and malicious activities are both acts differently.

3.1 Problemstatement:

A single compromised machine with its effective transmission capacity to launch various attacks can violate entire network in short time. Different kinds of attacks include denial of service attacks, stealing user identities, malware and spamming. With the existence of the large number of affected terminals, it has become a significant problem for the identification to the network administrators. Spamming is one of the major attacks that accumulate the large number of compromised machines by sending unwanted messages, viruses and phishing through emails.

A major goal of this research is to detect the machines that are involved in the spam communications in a real time network. The machines in a network involving in sending a spam are called spam zombies. A lot of research has been carried out to analyze the characteristics and behavior of machines in anetwork.

Previous studies like spam signatures, botnet analysis and spam killer has been carried out to study the detection of fraud machines in a network. Spam signatures contributed to analyze spam involved machines through URL signatures and succeeded in black listing in small networks. Spam killer is one of the tools that resolve the filtering of spam, to some extent, but much focused on the emails that are redirected to inbox.Itdoes not involve in identifying the client or spam details in the small/large networks. Botnet analysis and other studies analyzed the common characteristics and activities of bots which further helped to find geographic distribution of botnets[17].But still research has to be done as the previous studies are applicable only in small networks and lacks in bringing the global characteristics in an online manner.

3.2 Objectives

The objective of this research is to design and implement a tool to detect the spam attacks in a network.

Other objectives:

- To critically review spam characteristics in a network and analyze the current trends in protecting the spam inside a network.
- To identify the loop holes in the existing systems and propose an efficient tool.
- To design and develop proposed tool that identifies spam communications and keeps track of IP addresses of the machines in a network.
- To maintain the database of the monitored systems, it's time stamps, data and IP addresses using SQL.

- To run the test cases for various functionalities of an application in a real time.
- To evaluate the performance of the tool based on the amount of spam involved in the network, it's detecting capacity and its advantages over existingsystems.
- **FILTER TOOLDESIGN**

4.1 Proposed system:

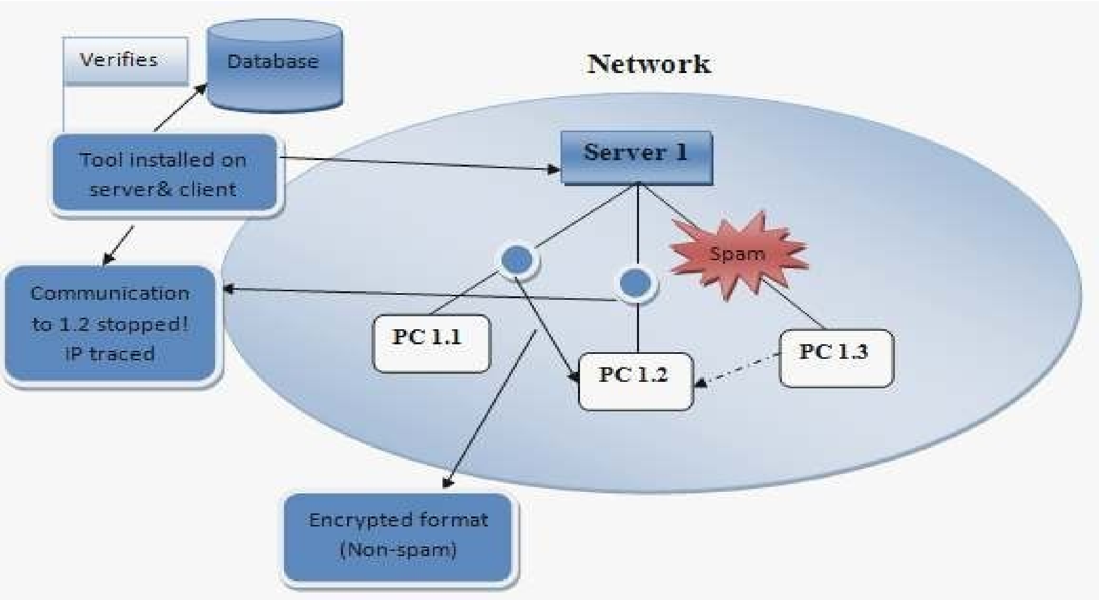

Verifies
Database
Network
Tool installed on server& client
Server 1
Spam
Communication to 1.2 stopped! IP traced
PC 1.1
PC 1.2
PC 1.3
Encrypted format (Non-spam)

Figure 3: System architecture

This tool is really useful in any organization in which servers and clients are connected in a network.Inreference to Figure 3, the clients (PC 1.1, PC 1.2 etc.) can communicate either in a network or even outside the network. All outgoing messages and internal messages sent by any client to other machines are routed through their individual servers (server1). The server scans the email and detects the spam that is being sentbythe client. The server receives the email and delivers to its corresponding destination nodes if the email is free ofspam.

The server can handle any number of clients in a network. It includes two sets of software, one for the server and the other one for the client. A client can sende mailto any emailID(Gmail,Hotmail,Yahooetc.).The client uses its mailboxwith the client software for sending any emails. The server software detects the spam based on the content spam detection [20] and differentiates between the spam email and the normal email.

4.2 **Spam filter process**:

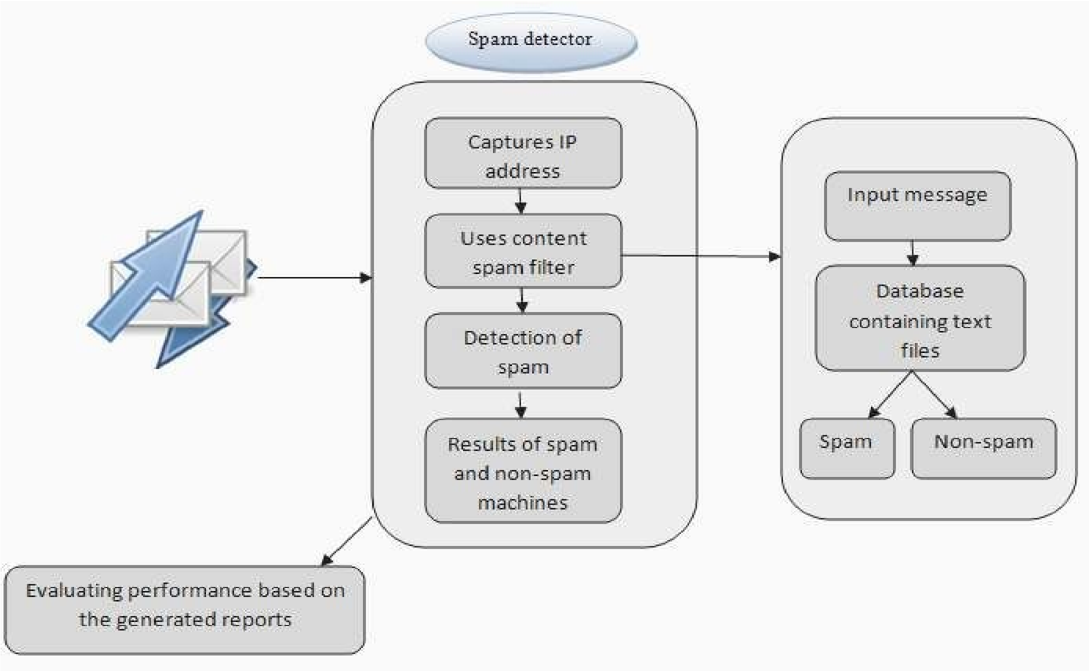

Spam detector
Captures IP address
Uses content spam filter
Detection of spam
Results of spam and non-spam machines
Input message
Database containing text files
Spam
Non-spam
Evaluating performance based on the generated reports

Figure 4: Spam filter process

Figure 4 depicts the system architecture of the spam detector that is designed and developed. The flow of execution that tool follows is also represented in the above architecture. The content spam filter takes the input text message and verifies.

4.3 Functionalities:

The flow of execution is explained in terms of different modules. This project is implemented in .net frame work. The back end is developed in Oracle SQL. All the modules and the transferring and receiving of data is explained clearly in this section.

4.3.1 MasterServer:

Master server is the component that is used by the network administrator to keep track of clients and spam details. This one has its private credentials for the network administrator to login into this application. It has number of functionalities like adding clients, sending data, receiving data and generating spam reports. [18]

4.3.1.1 Add clients:

Master server can add any number of servers and clients. Once the server is selected, any number of clients can be added to their respective server based on itsIPaddress. It has the options of adding client address based on theIPaddress. AnIPaddress is also validated when it is specified. Every single client is provided with its login credentials as well. User ID and password are assigned for all the clients to maintain personal mail boxes to send an email. When the client is added, its IPaddress, client name and its corresponding server gets listed in to the clients list.

4.3.1.2 Add server:

Master server has the functionality of adding any number of servers. Any number of clients can be added under these respective servers. Servers are added based on the systems IP address. It keeps track of the number of clients fall under each server. As there might be network traffic due to the number of clients in a single server, it has the option of adding additional servers. This reduces the bandwidth and network traffic across the network. Master server lists all the servers with its respective IP address when the server is added.

4.3.1.3 Spam report:

This is a very important component that is set by the master server where all the spam reports are maintained. Based on the content of the emails that are exchanged between the clients, the spam is recorded and there bya spam report is generated. It filters out the compromised machines and non-compromised machines based on the spam it is sent. The fields that are used by the master server in the spam report are client name, server name, IP address and time stamps. These fields give the clear picture of the clients that are connected to its corresponding servers.

4.3.2 Clientmodule:

The functionalities of client modules are clearly explained in this section. As discussed in the above section, the server is capable enough to handle any number of clients. The clients exchange the data and communicate with each other based on the following activities.

4.3.2.1 Send data:

For every client, there exist mail boxes through which data can be sent. The exchange of messages takes place in the form of an email.Ithas the options of composing an email through which information is passed on from one client to other clients. As this application deals with a content spam filter, email has the option of sending only text messages. It does not have any option of attaching a file as this may lead to transfer an image file, and spam cannot be detected if it is hidden in the image. A client can send an email based on the client's unique ID. Once and email is sent; all the information such as client's name, Ip address, date and time stamps are stored in the database.

4.3.2.2 Receive data:

When a client transfers an email in form of a text, it reaches to the inbox of another client machine. The mail boxes of clients are similar to that of the general email set ups and have an option to compose, inbox and sent mail. The client has also the feature to forward and delete the emails. One thing to be observed here is; if the email content represents spam, subsequently the tool blocks the message to deliver to the target client machine. If the email is a non-spam, then the message gets exchanged between the clients.

4.3.3 Spam reportmodule:

This is the most important module in this project. This module runs under the server software. It maintains the record of all the spam and non-spam messages, thereby identifying compromised and non-compromised machines in a network. It keeps track of machines according to the dates, server names and client names. It has the separate records like spam machines, non-spam machines and all together containing the client's name, IP address, spam details, date and time stamps. It also has the option of retrieving the records for a particular period. The spam is detected based on spam filtering algorithm. A major thing that is to be noted here is, spam details are reported whereas the non-spam report, and the messages are encrypted for the network administrator. This is because to maintain the privacy between the clients in a network.

4.4 Spam filteringalgorithm:

One of the most critical problems in today's world is the amount of spam emails spreading across the network. With the increase in day-to-day technology, the count of spam emails is getting into proportion. Spammers are also becoming effective as they tend to adapt to new techniques. This spam filtering based on the idea of Bayesian spam filters [13] can be considered as one of the most powerful techniques to identify and control the spam emails.

The legitimate and illegitimate spam emails can be well differentiated using this algorithm. The probability of occurrence of spam in the future is based on the past occurrences of spam. If any part of the email occurs frequently in spam emails but not in the normal email, then that part is more prone to spam identification. This algorithm verifies the content of an email against the information exchanged in the organization. Based on this information the bounding knowledge of the spam is filtered. The word that is used in an email can be a spam in one organization and may not be a spam in another organization. This algorithm verifies this particular word with the number of times it has been occurred with respect to organization and recognizes the spam based on the probability ratio. The important part of this algorithm is, it computes the detection of spam based on the patterns [14].It identifies the block of words against its spamicity instead of individual words for more effective detection of spam content in an email.

Advantages:

- This algorithm has the feature of implementing in any organization accordingly to the user. Though it takes a couple of weeks to analyze the email habits, it can distinguish well and gets updated with the new spam techniques.

- It considers a complete message instead of single words with respect to its organization. It can be referred as the intelligent approach due to its message examining criteria.
- It provides sensitivity to theclient and adapts well to the future spam techniques. Even if the spam word is slightly modified, this algorithm still succeeds and notices the spam content.
- This algorithm can be used as a multi-lingual approach due to its nature of being adaptive.

4.5 Encoding Technique:

The encryption for the non-spam email among the clients from the view of the network administrator is based on the Encoding. Unicode property. There are various in built encoding properties set in .NET. Some of them are ASCII, UTF-8, UTF-16 and UC- 2, UTF-7, Windows/ANSI code pages and ISO-8859-1. The GetByteCount method is used to find the count of bytes generated in the encoding a block of Unicode characters. The main part of encoding is done based on the GetBytes method. Similarly, GetCharCount method is used to find the count of characters generated in decoding a setof bytes, whereas the methods GetChars and GetString are used to execute the real decoding part.

Public static Encoding Unicode { get; } results in a Unicode format in little- endian byte order. The method System.Text.Encoding.GetBytes(System Char[]) is used to regenerate character arrays into arrays of bytes. This Unicode property provides an output encoding for the UTF-16 format using the same little endian byte order. The most commonly used method to save the Unicode characters on Intel computers are by using the native byte order. As an alternative, this property uses an array of bytes followed by the sequence of bytes for the purpose of encryption.

4.6 Classdiagram:

These diagrams are considered as the static structure of an application. The class diagrams include member variables and member functions or methods with the relationship among them. In the Figure 5, member variables and methods are clearly specified for each module.

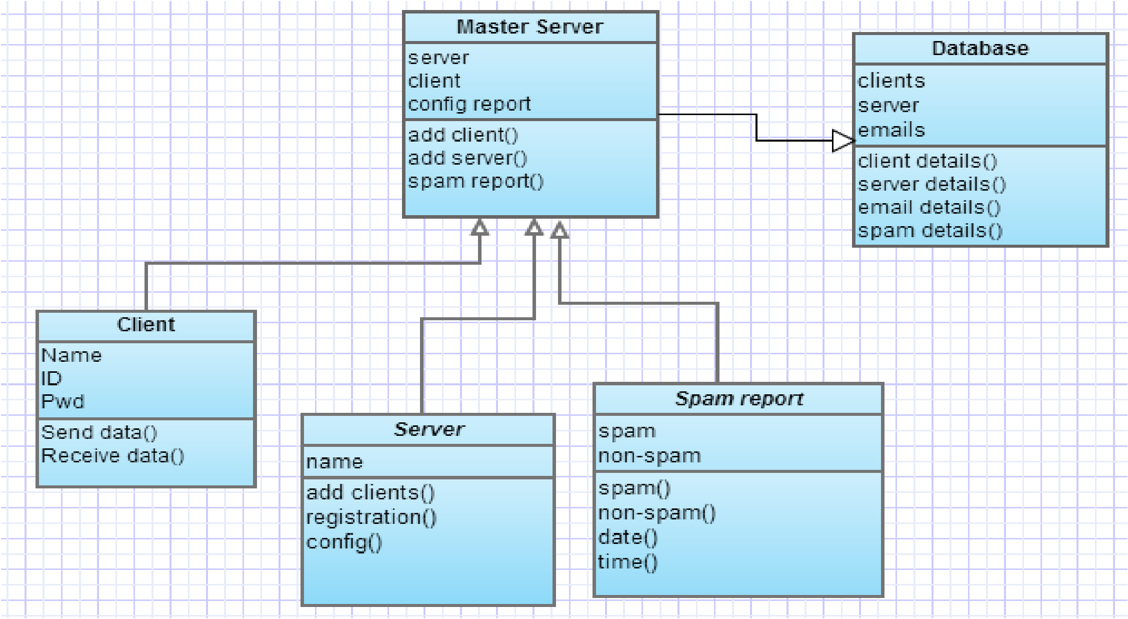


Figure 5: Class diagram

Use case diagram:

Use case diagrams typically define the use of a system. It depicts the relationship and communication between the actors and a system to reach the targets of a system. Any external machine or a human can be applicable as an actor. In the Figure 6 below, the activities between the master server and among the clients are clearly represented.

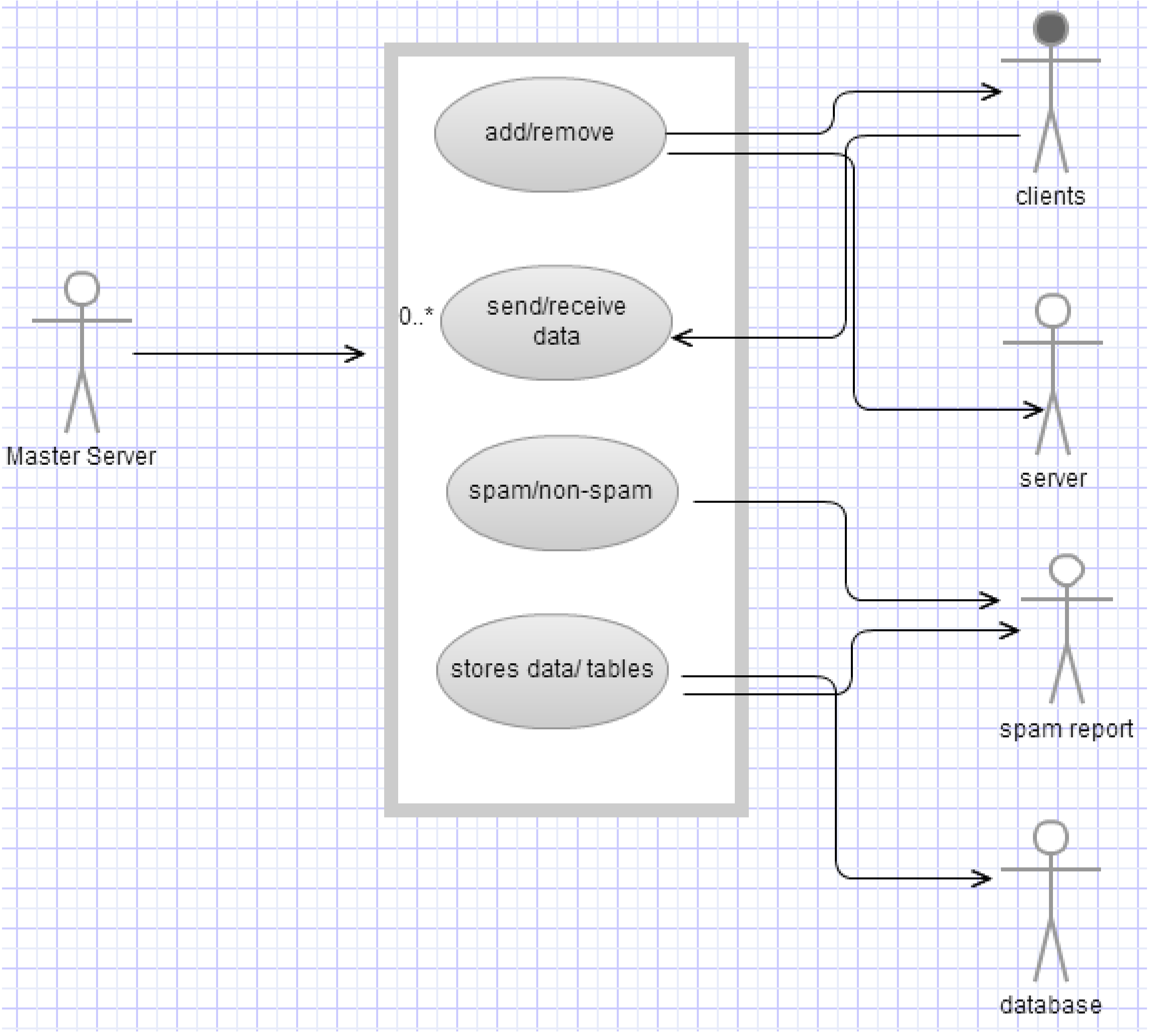


Figure 6: Use-case diagram

Sequencediagram:

Sequence diagrams can also be referred as even diagram or timing diagrams. It typically defines the processes including in a system, how they interact with each other and its priorities. The object interactions perform in a specified sequence. It also provides a clear picture of objects and classes along with ordering of messages among the objects for specific function. Figure 7 clearly represents the exchange of messages and activities between the modules.

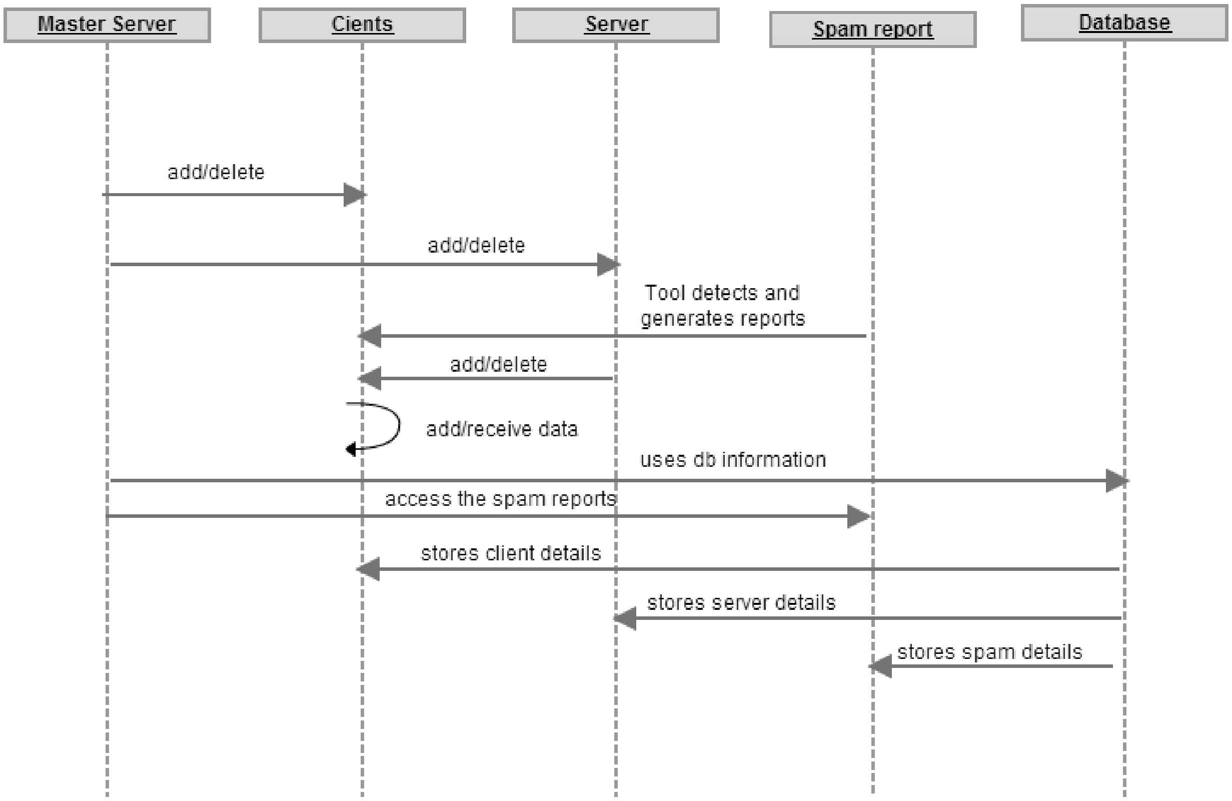
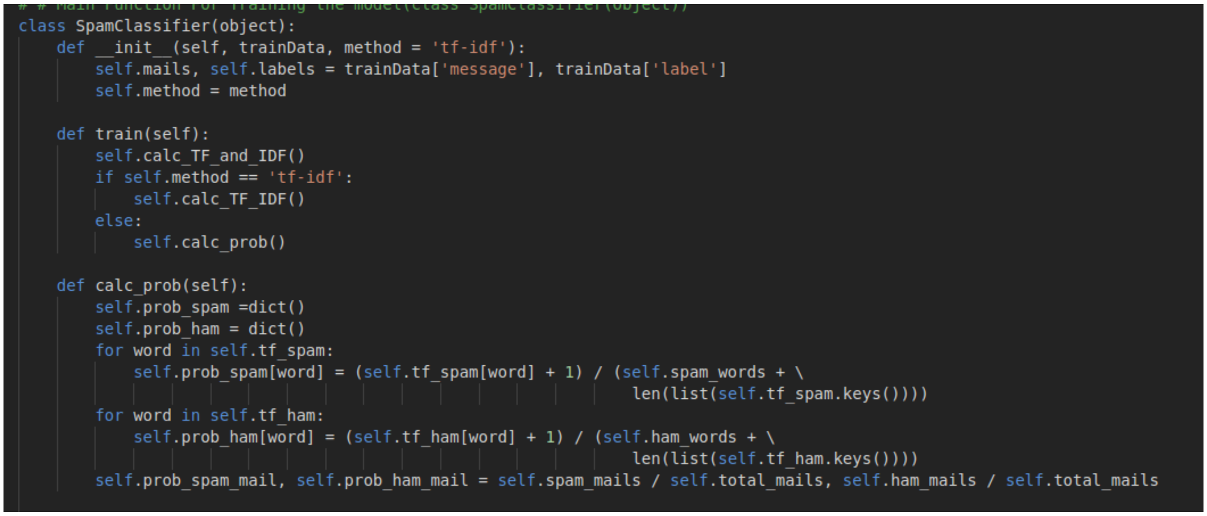


Figure 7: Sequence diagram

IMPLEMENTATION

Modules:

7. **class Spam_Classifier Object:** It is basically used to divide the keywords in terms of list so that the model can be used to find the repeatence of the word in the list and find out is there is any need to consider the word in the group or not.

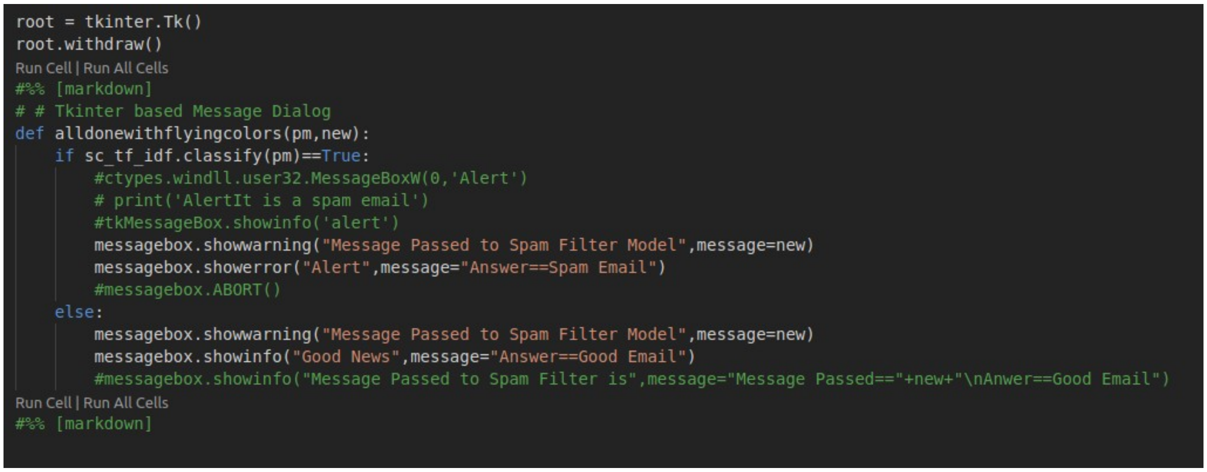


```
# # Main Function for Training the model(class SpamClassifier(object))
class SpamClassifier(object):
    def __init__(self, trainData, method = 'tf-idf'):
        self.mails, self.labels = trainData['message'], trainData['label']
        self.method = method

    def train(self):
        self.calc_TF_and_IDF()
        if self.method == 'tf-idf':
            self.calc_TF_IDF()
        else:
            self.calc_prob()

    def calc_prob(self):
        self.prob_spam =dict()
        self.prob_ham = dict()
        for word in self.tf_spam:
            self.prob_spam[word] = (self.tf_spam[word] + 1) / (self.spam_words + \
                                                               len(list(self.tf_spam.keys())))
        for word in self.tf_ham:
            self.prob_ham[word] = (self.tf_ham[word] + 1) / (self.ham_words + \
                                                               len(list(self.tf_ham.keys())))
        self.prob_spam_mail, self.prob_ham_mail = self.spam_mails / self.total_mails, self.ham_mails / self.total_mails
```

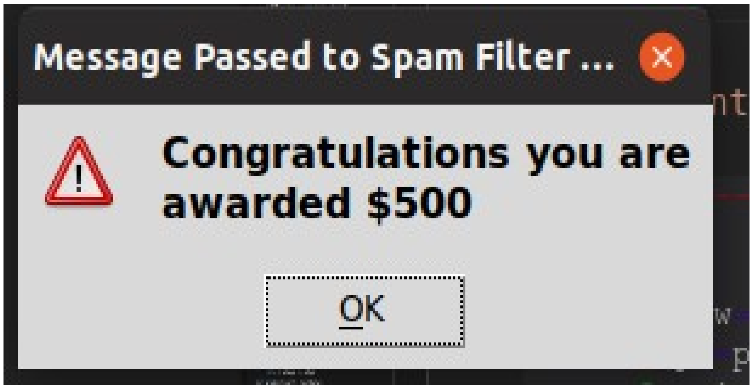
Tkinter box: It is used to show the message in the result from the def alldonewithflyingcolors(pm,new) is going to show the result through the text box.



```
root = tkinter.Tk()
root.withdraw()
Run Cell | Run All Cells
#%% [markdown]
# # Tkinter based Message Dialog
def alldonewithflyingcolors(pm,new):
    if sc_tf_idf.classify(pm)==True:
        #ctypes.windll.user32.MessageBoxW(0,'Alert')
        # print('AlertIt is a spam email')
        #tkMessageBox.showinfo('alert')
        messagebox.showwarning("Message Passed to Spam Filter Model",message=new)
        messagebox.showerror("Alert",message="Answer==Spam Email")
        #messagebox.ABORT()
    else:
        messagebox.showwarning("Message Passed to Spam Filter Model",message=new)
        messagebox.showinfo("Good News",message="Answer==Good Email")
        #messagebox.showinfo("Message Passed to Spam Filter is",message="Message Passed=="+new+"\nAnwer==Good Email")
Run Cell | Run All Cells
#%% [markdown]
```

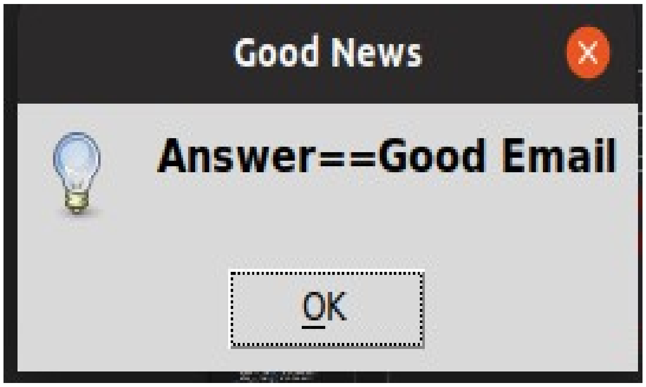
process_message(new)== new is the string which used to take the string in it and send it to the Spamclassifierobject so that it can further process it.

SCREENSHOTS OF RUNNING



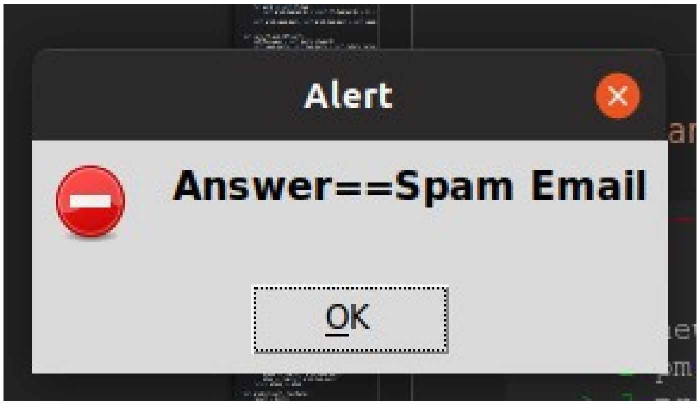
Message Passed to Spam Filter ...

Congratulations you are awarded $500



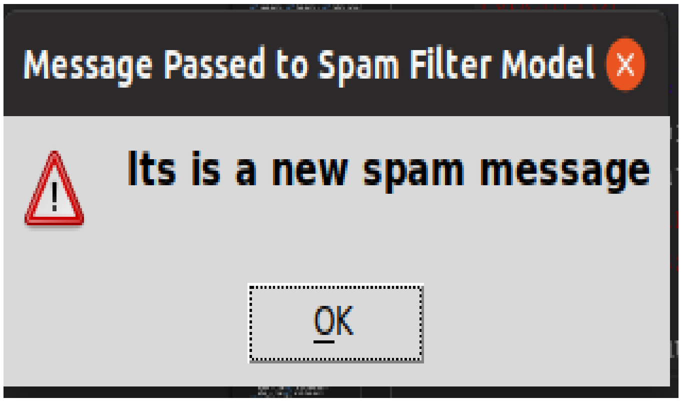
Good News

Answer==Good Email



Alert

Answer==Spam Email



Message Passed to Spam Filter Model

Its is a new spam message

Testing

Testing is a very important module in the software development to verify, validate and provide quality and service for different components of software. It is used to minimize the risks by efficient use of resources in the development life cycle. This module can be employed at any point of the development process. It is efficient for the testing phase to be implemented at initial level to lower down the risks of defects and failures.

CONCLUSION & FUTURE WORK

Due to enormous usage of internet technology, there is a huge increase in the network attacks. Among them, spam is considered as one of the main attacks in launching various attacks like stealing user identities and spreading malware etc. In this project, a spam detector is developed, which can monitor and detect the machines involved in spam across the network. This tool is based on a spam filtering algorithm that has the efficiency of detecting high percentage of spam. It can differentiate spam and non-spam affect machines in a network of any size. To avoid network administrator to view non-spam emails and to maintain the privacy among the clients in a network, encryption technique is used to encrypt them. The performance is evaluated based on the functionality and results generated with respect to the drawbacks of existing systems using the algorithm. This tool is considered as the light-weight tool because of its minimal amount of time and observations to detect a spam. It can also be used in a network consisting of any number of clients by providing an aggregate large-scale view of the spam in an online manner.

In future work, this tool can be extended to image spam detection as this one is completely based on the content spam filtering. It can be further enhanced by incorporating the sending message service feature to personal contact numbers if the spam exceeds the assumed threshold value. And finally, apart from spam attacks several other attacks can also be focused along with the protective measures

REFERENCES

1. **Building a Spam Filter from Scratch Using Machine Learning—Machine Learning Easy and Fun:** https://medium.com/analytics-vidhya/building-a-spam-filter-from-scratch-using-machine-learning-fc58b178ea56
2. **Github Link for the Project:** https://github.com/abhijeet3922/Mail-Spam-Filtering
3. **Buid a Spam Filter Blog from Machine Learning Expert Yuva:** https://www.pythonforengineers.com/build-a-spam-filter/
4. **CS430 Class Spam Filter**
5. **Spam Filter Blog from the KD Nuggets(A very interactive blog):** https://www.kdnuggets.com/2017/03/email-spam-filtering-an-implementation-with-python-and-scikit-learn.html