

## Software Agent for E-mail Spam Filtering

**Laith R. Flaih**

*Computer Science Department  
Cihan University – Erbil  
Erbil, Iraq.*

**Mohammed Hussein Shukur**

*Computer Science Department  
Cihan University – Erbil  
Erbil, Iraq.*

### Abstract

With wide acceptance of personal computers, many people spoke of a social change from industry to information. Years ago some people predicted the amazing growth and newer universal acceptance of an electronic system for sharing and exchanging information then the world witnessed a communication reevaluation initiated by the internet that bridged time, distance and culture. One of the most important services that the internet provides to its users is the Electronic Mails (E-mails). The main objective of this paper is to design and implement a complete message switching and filtering system that solves the spam or junk mail problem using software agents and rule –based system, the proposed system able to check each incoming and outgoing E-mail to blocks the unwanted messages or replaces specific undesired words from other message before being sent or received, the system offers the creation of dynamic number of agents according to user desire each agent can have dynamic number of filtering rules that satisfies the requirements.

**Keywords:** Agent, Rule-Based System, E-mail System, Spam Filtering.

### INTRODUCTION

The most popular way that is used to access and use internet services is the World Wide Web (WWW). The reason that the WWW is popular and widely used by many people all around the world is that the web allows full, high quality color graphics, sound, videos and texts that make it an attractive multimedia tool. Furthermore, all the resources that are available throughout the net, the ease of using and accessing those services and resources and its cheapness compared with other information sources and communication means, made internet popular with millions of computer users around the world[].

One of the most important services that internet provides to its users is the Electronic Mail (E-mails) Actually E- mail service is older than the internet itself, it was just advance on what we know these days as file directory. It is represented by just placing a message in another user directory in a post where they can see when they log in to their systems. An example of the very fast E-mail system of such kind are MAILBOX which was used at Massachusetts Institute of Technology in 1965 and SNDMSG mail system which was used to send messages on the same computer[2].

In 1972, the symbol “@” was used to denote sending a message from one computer to another. Message switching service has developed rapidly after that until it become one of the most important and popular internet services due to its speed, low cost, availability , offering the offline mode messages and it reaches all-around the world where ever the internet service exists. Nowadays, it is not only used by millions but hundreds of millions of people around the world [3].

### PROBLEM BACKGROUND

With the growing use of this important internet service (E-mail), people started using it not only for exchanging important information, but also for other purposes, for instance companies and manufacture owners started using it to advertise for their products and services, and other started using it for malicious purposes and they used E-mail to deliver viruses and harmful contents to others accounts, such content may cause a whole system crash which finally may cause severe losses could reach to millions of dollars. It is also used by abusers to send content that are no appropriate for kids or under aged persons.

Unsolicited Bulk E-mail (UBM), Unsolicited Commercial E-mail (UCM), spam, bulk or junk mails are all terms used to indicate the unwanted messages arrives mail boxes in a daily time fashion without the approval of it recipients. Such mails are annoying not only to its recipients but also to the internet service providers because they occupy much bandwidth, wasting many resources of computation and storage, threatening the safety of internet and personal computer and it does not only frustrating most users, it strains the IT infrastructures of organization and cost businesses billions of dollars in lost productivity [4].

The number of spam mails increased and started to be considered as a real problem especially when the number of mails in the user account mail boxes have been increased exponentially, where in 1998 Cranon and Lamacchia found that 10 % of incoming E-mail to the network was spam, in 2000 World Talk ZCrop denoted that over 60 million business people have been overwhelmed since 30% of total E-mail were spam, while Eu Gene Kaspersky in 2004besimated the percentage of spam mails about 70% of total incoming E-mail[5], in 2010 the percentage raised to 89% according to Internet Security Threat Report(2013), because of the growing popularity of social means like Twitter, Facebook and others, spammers moved their activity of spreading spam mails through these media[6].

## E-MAIL SYSTEM

E-mail system was first designed to exchange messages on the same machine until 1972 when the first E-mail message was sent between two remote devices by an engineer named Ray Tomlinson who used the “@” symbol to designate the receiving machine[1].

The e- mail system is mainly compose of two parts [7]:-

- A. E-mail Server is a host that delivers, routs, and stores E-mail messages as a text file for each account, this text file stores information like the sender address, receiver address, date and time of recipient plus the E-mail subject line and message body formatted and appended to the bottom of the text file.
- B. E-mail Clients, this acts as an interface between the user and E-mail server, it connects with the E-mail server and allows the user to access the logged in users text file on the server and perform one of these actions: listing the incoming E-mails, reading the highlighted and selected message header, composing new E-mail message and adding an attachment files to the sent E-mail.

## SPAM

No existence of exact definition for the term spam, lead us to look for a general agreement that spam is “unwanted E-mail”, but must clarify that not all unwanted messages are spam. Because spams are different from every user, there sre more than one term used to denote to this type of E-mails, terms like, Unsolicited Commercial Mails (UCM), Unsolicited Bulk Mails (UBM), Junk or Bulk mails are all used to denote to messages sent in bulk to large group of recipients [8].

The first E-mail list offered for sale with more than 2 million addresses then it started to become a business. In 1997 spamming problem become totally out of control and at the end of the same year, number of spam mails showed an exponential growth [8].

After this period of time the urgent need for spam filters rose and the big breakthrough was in 2002 when machine learning and statistical classification techniques were first introduced by Paul Graham in his famous article “A Plan for Spam”, he wrote in his article” I think it is possible to stop spam, and that content-based filters are the way to do it “, [1].

Since that time and until nowadays the need for more and more reliable and efficient spam filters become substantial because the spam rate reached irritating rate considering its cheap coast and illegal form of advertisement that reaches millions of users connected to the internet also lead to a never ending fight between spam filter programmers and spammers who always find some way through these filters [9].

## THE PROPOSED SYSTEM

The system proposed in this paper is a complete message exchanging and spam filtering system that use filtering software agent that is subclass of task specific agent. The

software filtering agents are based on rule-based system to impact each incoming and outgoing message throughout the system and as a result the incoming/outgoing messages will either be blocked, filtered or sent/received as any legitimate message in the user inbox. The system.

The system offers building and adding new agent to the system without the need to program the new agent, it simply can be defined any user who has or has not programming skills.

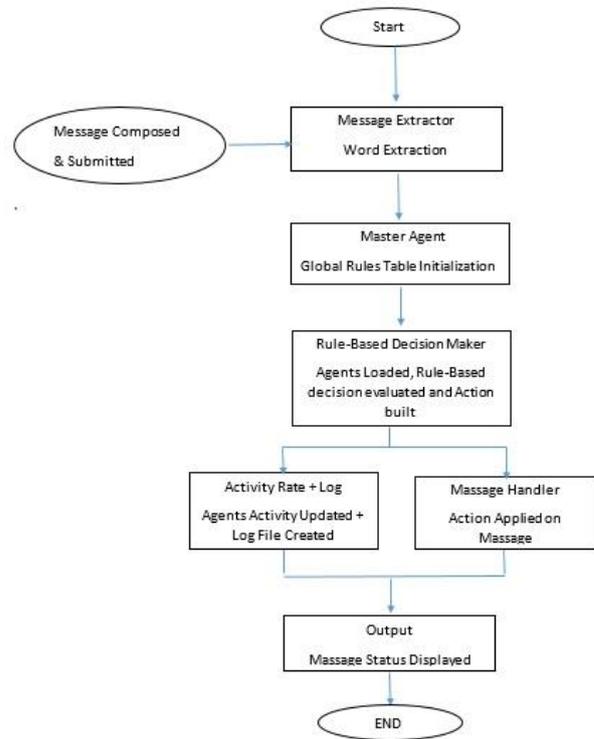


Figure 1 show the main components of the proposed system

After the message has been composed by the user and submitted to the system. Message Extractor will handle the task of message decomposing and analyzing, it checks the submitted message starting from its subject and identifies its beginning and the end of the message then the message will be read word by word, the Message Extractor starts word extraction and retrieves table or array contain the set of words that the message is composed of.

The extracted words and the list of agents which consists of the set of all agents will be delivered to the Master Agent which in turn initializes the global rules and according to these rules the master agent will perform the task of distributing the incoming message to the most appropriate available agents after determining active agent from the non-ones (as the agent can set to be active or non-active, and it is up to the user). The user can invoke or use more than one agent at a time.

After that the Master Agent has determined the active agent set, organized and distributed the incoming E-mails, the Rule-Based Decision Maker initialize the local rules for each agent which represents the core of each agent, when an agent receive

an incoming E-mail from the master agent it depends on these rules to decide whether to block, filter out (replacing specific undesired words) or simply pass it to its destination. This action will be done by matching the extracted word table against these local rule if rules match or simply if the message contains specific block or undesired words then appropriate action will be performed on that message.

When the Rule-Based Decision Maker finishes its task and the action to be performed on the delivered E-mail has been decided by the chosen agent, the activity rate for that agent will be increased with each block or word replacing action, then a log file will be created, the file size and content will be update continuously with each taken action by that agent and the new replaced or blocked word will be added to the file which in result will decrease the time needed for decision making in future.

The final component which is the Message Handler will specify the action applied on the message whether blocking or filtering, According to the relevant agent's rules, the message will either be blocked or specific words will be replaced and message status will be displayed as output.

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the file "MSW\_USltr\_format".

## CONCLUSIONS

During the implementation and testing phase of the proposed system, set of conclusions have been drawn, these conclusions are listed below:-

- The system succeeded in achieving the goals which designed for, all messages are filtered correctly.
- The system is easily to use. The designed GUI facilitates the interaction and communication between the system and users.
- The proposed system is one of the filtering systems

that offer creation of new filtering agents without re-programming or changing the code of existed agent codes in very simple and interactive way.

- The capability to view filtered message after filtering process concealing inappropriate words from the filtered messages is another feature that characterizes the system and distinguishes it from other filtering systems.

## REFERENCES

- [1] Abdullah, S., 2002, "Internet and Arabic Search Engines", MSc Thesis, Al-Nahreen University.
- [2] Brown, B., 2007, "The Complete Guide to E-mail Marketing", USA, Atlantic Publishing Group.
- [3] Mahdi, A., 2007, "Using Mobile Agent for Development of Web Site Search Engine", PhD Thesis, Information Institute for Postgraduate Studies of the Iraqi Commission for Computer and Informatics.
- [4] Christina, V. et al, 2010, "A Study on E-mail Spam Filtering Techniques", International Journal of Computer Applications", Vol.12, Issue 1.
- [5] Symantec Corporation. Internet Security Threat Report 2013, Vol.18.
- [6] Lazzari, L. et al., 2005, "CAFÉ-Collaborative Agent for Filtering E-mails", IEEE Computer Society Washington, USA.
- [7] Brain M., 2004, "How E-mail works", online, <http://www.armchairpatriot.com>.
- [8] Gulyas, C., 2006, "Creation of a Bayesian Network-Based Meta Spam Filter using the Analysis of Different Spam Filtering", MSc These, Mszakies University.
- [9] Wu, C., and Cheng, K., 2005, "Using Visual Features for Anti-Spam Filtering", IEEE CVPR.