A Novel Anti-Spam Solution to Minimize False Positive Alarm Rate in Bulk Spam

Abdulkareem Al- Alwani
Computer Science & Engineering Department
YUC, Royal Commission
Institute & Colleges
Yanbu, Saudi Arabia

ABSTRACT
Spam has emerged to be an impending problem for internet users and operators as the extensive penetration of spam is now affecting almost every entity associated with the internet. Email filtering and blocking efforts by network operators, software vendors and Internet service providers (ISPs) are the key actions that are used to stop spam before it reaches their users. Recent estimates by reliable organization nonetheless indicate that spam makes up between 70% and 80% of email traffic worldwide. Thus, spam can create a significant burden for network operators, and the problems associated with spam are more dominant in developing countries, where high volumes of incoming and outgoing spam can cause a severe drain on the already choked bandwidth that is available in those regions. While providing an overview of the impact caused by spam and the efforts carried out worldwide to fight spam, a novel solution for spam analysis is presented in this study. The main strength of the proposed approach lies in the reduced rate of false positive alarm, i.e. detection and rejection of valid emails. Five functional blocks along with six spam analysis techniques were employed in this anti-spam solution which operated in a sequential manner with intelligently designed rules to minimize false positive emails. The proposed model was tested on 56 email accounts handled by enrolled students in three major departments of a local Saudi Arabian University over a span of one year and results showed a marked decrease in false positive alarm rate. The model proved to be an effective solution for multi-tier anti-spam defense and can be employed in various applications to mitigate spam outreach in online email and messaging services.

General Terms
Anti-Spam, Bulk Spam

Keywords
Spam, Anti-Spam, Anti-Spam laws, Spam Email, Spam impact, Social networking; spam; anti-spam strategies

1. INTRODUCTION
In recent years, the evolution of information technology has revolutionized the way humans used to interact among themselves. Entities from different parts of the world can interact and share, freely and seamlessly at lightning speeds only linked by the ever growing internet system. The World Wide Web and the internet now have thousands of applications and services available for users to use and communicate. Despite these numerous applications, EMAIL is still the most used and reliable tool for effective communication and correspondence [1].

Development of the Internet email technology took more than 20 years, and it is still under progress. The electronic mail (email) has been one of the biggest technological and sociological development in the history of internet. As EMAIL is one of the most used online services for correspondence and communication on the internet; the users of EMAIL are more prone to security vulnerabilities [2]. Any single weak link in the EMAIL tree of the user cannot only expose email account of a single individual, but it will expose every individual's account in that tree. Spamming is the most common form of exploitation and it is extensively used to take advantage of vulnerabilities of in an EMAIL system. Email spam is junk or unwanted email and it is one of the most annoying and invasive problem faced by users of Internet. It not only brings financial damage but also affect users to personal level.

A recent study on spam activities showed that over 70% of the commercial and business emails are spam [3]; therefore, there are severe issues related with increasing volume of spam such as overloaded mailboxes, haphazard personal mail, storage space problems and saturation of servers. Wastage of time is another major issue that is caused by spam as it time consuming to filter and delete spam emails.

Spam mails vary significantly in their purpose and they roughly belong to the following categories: forgery, scams, and shortcuts to money making, miracle medicines, business forgeries, pornography invites, fake friend requests, and advertisements. [4].

The fact is that spam or junk email is an overwhelmingly offensive method of email marketing. The motive behind spam is mostly of commercial nature, and spam is always carried out in large numbers and repeatedly. Therefore a uniform and well established approach is required to address spam on all fronts.

Next section will provide a brief overview of the impact of spam activities, anti-spam laws and regulations implemented worldwide and the challenges faced in implementing anti-spam measures..

2. RELATED WORK
Email is now a global tool to communicate with customers, employees, family and friends. As email can be sent and received freely and through a click of few buttons, it has also become the means of fake emails and invasive advertisements of financial services, pornography, piracy and drugs. A large percentage of emails are spam, which shows how resources are being wasted in handling spam email.

Spam has quietly become a common feature of our email accounts as accounts become overloaded due to unwanted bulk spam, making the account unmanageable and time consuming. The situation is such that bulk volume of spam

emails delivering to mail transfer agents is equivalent to the effect of denial.

Spam is counterproductive in every aspect and wastes systems and people resources [5]. Spam decimates disk space, bandwidth and makes it very annoying for an individual to manage his email account and prevents reception of important emails as eventually no space will be left in the email account [6].

The purpose of reviewing literature in this study is to get accustomed to the basic impact of spam and the laws that are being implemented to stop spam activity. As the laws and legislations define the nature of the target and the spam, the effectiveness of these laws can be exponentially increased using a viable anti-spam solution which has the essential components required to implement effective anti-spam measures.

2.1 Impact of Spam

Spam emails and messages, as simple and innocent as they seem, have far reaching effects and consequences. Following list summarizes the impact caused by spam and related activities.

- Spam adversely affects the results of search queries and limits revenue of legitimate websites.
- Spam at the backend influences online entities financially and this can be used to undermine open competition between vendors.
- Spam weakens the trust of a user on an online system or service he is using. In turn the system or email service becomes a victim for user's reaction to spam; despite not being involved in spamming, it is thought to be responsible by the end user. This gives rise to trust issues at the user side.
- Spam can be disseminated vide websites, chat rooms, online forums where money, porn and lottery are some prime examples of enticements used to lure users.
- Spam forces a user to revisit his account repeatedly wasting time, and when an account is spammed, there is always a possibility that this email address is now available for other spammers too.
- Emotional and mental distress can also become pronounced as online lottery and gift giveaways are always getting normal users to waste their money on scams, and in return they get nothing.
- Spam destroys privacy as an email address shared with an unknown entity online will definitely be comprised again and again.

Spam emails pose a serious technological and economic challenge. Numerous techniques are therefore applied to limit the volume of spam email as much as possible. These techniques can be broadly categorized in to two groups which are,

a) Boundary based: This approach establishes certain rules to curtail the spam activity e.g. by allowing access or authentication of being a human using a CAPTCHA or by placing usage limits e.g. Flickr placed a limit of 75 tags for a single photo to stop spam tagging) [7].

b) Classification based: These approaches use classification and statistical tools to identify spam mail. Actions can vary from blocking spam to deleting the spam automatically. In detection based approaches, spam can be considered as an object with certain attributes and these attributes helps in the classification of unwanted emails from authentic emails [7].

2.2 Anti-Spam Laws and Regulations

The potential of these strategies along with other approaches can be fully utilized, given a platform and regulation is provided for implementation backed by governments or regulating bodies.

Countries and organizations around the world are taking serious steps to fight spam, although these efforts are often carried out in developed countries. Countries around the world are developing legislation to improve anti-spam measures. Major legal initiatives taken by different countries are as under:

2.2.1 Australian Spam Act and Codes of Practice (2003):

This law comprehensively regulates email and message based communication intended for commercial purposes. As per this act, it is illegal to send unsolicited online messages of commercial nature with origin in Australia. Offenders can be fined with a penalty up to $1.1 million a day for repeat corporate offences [8].

2.2.2 Canadian Anti-Spam Act (2010):

The Canadian legislation [9] defines the legal requirements of circulating commercial email along with penalties for originators of unsolicited emails. To ensure active implementation of this legislation, a Spam Reporting Centre (SRC) was developed where complaints related to unsolicited commercial electronic messages can be registered.

2.2.3 European Commission e-Privacy and Electronic Communications Directive (2002):

This directive was issued in 2002 and it asks Member States to bar sending of unsolicited commercial emails and messages unless the prior consent of the destinations has been registered [10].

2.2.4 US CAN-Spam Act (2003):

This act proposes a legal framework to prevent spam and allows users to opt-out at their own discretion. Most of states also have their respective laws have enacted to control spam activities [11].

Other examples of country based regulations and laws are: the New Zealand Unsolicited Electronic Messages Act [12], Japan's anti-spam law [13], the Singapore Spam Control Bill [14].

Similarly, international collaboration for an anti-spam legislation has resulted in following developments:

2.2.5 OECD Anti-Spam Toolkit (2004):

In total, 34 countries volunteered for the development of OECD Anti-Spam toolkit [15]. This anti-spam toolkit is basically a regulatory handbook tracking best anti-spam practices, self-regulatory arrangements, a resource center for assisting users in self-protection against spam and an entity relationship inventory of spam origins.

2.2.6 APEC Principles for Action against Spam (2005)

In a declaration by Communication Ministers of the 21 Asia Pacific economies, a program of action along with governing

rules was proposed [16]. Implementation was set to be voluntary, similar to the OECD toolkit.

Prior to the announcement of this policy, sufficient technical homework was carried out along with policy level decisions in member countries leading to a conformance in regulating spam in these countries.

2.2.7 United Nations International Telecommunication Union (ITU) [17]:

Based on a directive from UN in 2004, the ITU started study groups to conduct research and develop methods for combating spam to at a global level. Since then, ITU has been an important resource and an active player in proposing techniques to counter spam and related threats. All these laws and regulations are reliable enough to resist influx of spam emails and allow normal email by deploying anti-spam approaches.

2.2.8 UN Economic Commission for Africa and the African Union

This commission is working on the draft Cyber Legislation in Africa (2012) [18]. The convention would address for key communication areas: cyber-security, data protection, e-transactions, and achieving conformance in legislation related to cybercrime across the region.

Broadcast or target spamming is an inexpensive and rude form of commercial marketing, directly or indirectly attacking people's privacy. It affects a large segment of public and costs are normally borne by the user. So, it always comes down to an individual to handle spam, as any strategy and regulation has its limitations. Next section will discuss the challenges pertinent to countering spam, effects that must be accounted for to stop spamming activities.

3. CHALLENGES IN COUNTERING SPAM

In Major regulation efforts to restrict the amount of spam activity will still take time to mature as there are numerous challenges which are being faced by organizations around the world in combating spam. Most impending ones are as follows of them are as follows:

a) In a trust based model for the prevention of SPAM, a user's trust can change overtime based on his experience in online networks. As trust changes dynamically over time for various services and entities, it is difficult to carry out a quantitative analysis of trust at an individual level all the time.

b) Majority of the anti-spam techniques use text based inference in a monolingual environment to identify spam. This limits the ability of the algorithm to attain optimum performance.

c) Once an email address is compromised, it is almost impossible to revert back the potential damages as incurred through spam. Only safe option is rolling back the email account but at the cost of losing a primary email account. Once an email account goes public, you cannot reverse the event as the address will remain public and open for spam attack.

d) Cross sharing of SPAM related data for conformance among different platforms has still not developed to ensure standardization in countering SPAM.

Statistical and probabilistic techniques are still to be complemented with machine learning to achieve a high rate of success in identification and subsequent classification of spam.

4. PROPOSED ANTI SPAM SOLUTION

An effective anti-spam framework must involve multiple layers of defense at different tiers of an organization. From the upper management level to the lower management level, the execution method of anti-spam techniques may differ. But as a whole, those anti-spam core components must be identified that are essential for an effective anti-spam solution. Following solution is proposed that constitute the main elements required to achieve conformance with the basic anti-spam laws implemented worldwide and to effectively prevent spam. A generic block diagram for the proposed solution is shown in Fig-1.

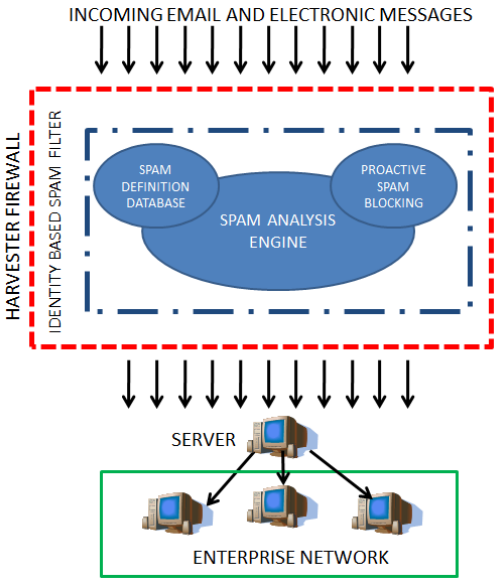

Fig 1: Proposed Anti-Spam Solution

4.1 Spam Analysis Engine:

An effective Anti-spam solution must include a Spam analysis tool that uses an intelligent technique to identify and analyze spam emails and messages. However, in this proposal, an analysis tool is proposed with a specific set of spam analysis techniques. Spam analysis engine in the proposed solution will incorporate multiple, overlapping anti-spam analysis techniques, and provide a mechanism for dynamically updating its spam definitions to maintain effectiveness against evolving spam activities.

Spam Analysis Engine is the core of the proposed anti-spam solution and Fig-2 shows the flow diagram of the proposed Spam analysis Engine. Major advantage that can be attributed to Spam analysis engine is that it can accommodate following prospective analysis techniques. These techniques can be applied individually or in combination with each other depending upon the requirement and nature of the enterprise, but all of these techniques are available for implementation at all times based on the choice of the user and spam threat level. A higher level flow diagram is shown in Fig.2.

Table 1 through 6 presents the results for all the above mentioned techniques for a total of three datasets. Emails for datasets were collected over a span of one year from a university in Saudi Arabia. A total of 56 email accounts belonging to students from three different departments were used. Humanities department dataset included 25 accounts and 31,800 emails, Physical science department included 19

accounts and 26,512 emails and engineering department included 15 accounts and 20,280 emails. All these accounts

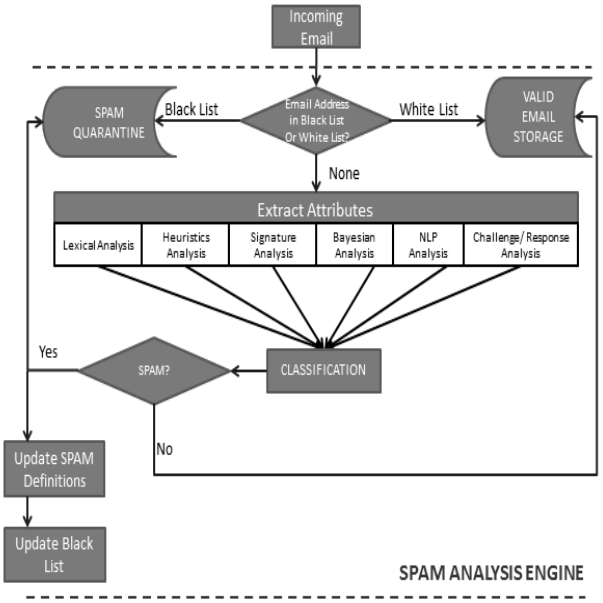

Fig2 :Spam Analysis Engine Employing Multiple Analysis Techniques

were exposed online on selected forums and websites at regular intervals throughout the year. Statistics were recorded for every technique individually and in combination. That is each email is checked for spam using each analysis technique sequentially, and if an analysis block raises a red flag, the email is tagged as spam with reference to that specific analysis technique id and is then pushed to next analysis block. Only emails which were rejected by only 2 or 1 analysis block were sent to Quarantine for checking false positive emails, rest were sent to spam or inbox of the recipient. Each email is tagged with 'SPAM' or 'NOT SPAM' with respective unique analysis technique ID.

4.1.1 Lexical Analysis

In the lexical analysis used in this solution [19], words and phrase lists based content filtering was carried out on each electronic message to identify scam. ID Tokens were generated against keywords, and these tokens were subsequently used for spam identification and classification. It was evident from results that inorder for lexical analysis to be successful; it should analyze at least message subject, body, attachments, and HTML tags, and also digs hidden or disguised text.

Table 1. Results for Lexical Analysis

DATASET	TOTAL MESSAGES	SPAM	NOT SPAM	FALSE POSITIVE ALARM
Humanities	31,800	73%	16%	11%
Physical Sciences	26,512	72%	15%	13%
Engineering	20,280	69%	17%	14%

4.1.2 Heuristics based Analysis

A classification based meta-heuristics approach was employed for this type of analysis [20]. Classification was carried out determine to whether an email message is spam or not.

Table 2.Results for Heuristics based Analysis

DATASET	TOTAL MESSAGES	SPAM	NOT SPAM	FALSE POSITIVE ALARM
Humanities	31,800	76%	14%	10 %
Physical Sciences	26,512	71%	13%	16 %
Engineering	20,280	71%	16%	13%

4.1.3 Signature Analysis

Signature analysis [21] was used to register 'hashes' from previously detected spam to build a library of threat signatures in spam definition database, and then it was used to correlate existing spam attributes with signatures in the incoming email. Matching entry that has a positive correlation was then marked as spam.

Table 3.Results for Signature Analysis

DATASET	TOTAL MESSAGES	SPAM	NOT SPAM	FALSE POSITIVE ALARM
Humanities	31,800	70%	11%	19%
Physical Sciences	26,512	71%	9%	20%
Engineering	20,280	78%	10%	12%

4.1.4 Bayesian Analysis

Bayesian analysis [22] is a commonly used technique. An improved Bayesian analysis model was incorporated which was trained using spam definition database to analyze aggregated textual attributes of an email. This technique is a proactive technique and it was selected for its better classification performance. This analysis classified spam based on probabilities of spam attributes previously found in emails.

Table 4.Results for Bayesian Analysis

DATASET	TOTAL MESSAGES	SPAM	NOT SPAM	FALSE POSITIVE ALARM
Humanities	31,800	75%	14%	11%
Physical Sciences	26,512	74%	12%	14%
Engineering	20,280	77%	11%	12%

4.1.5 Natural Language Processing based Analysis

NLP based technique [23] was employed in combination with machine learning to assess and analyze text and correlate it based on morphemic, syntactic, and pragmatic analysis to extract meanings pointing to spam content. This is another prominent technique which is commonly used for proactive filtering of spam emails and recognizing bots.

Table 5. Results for NLP Analysis

DATASET	TOTAL MESSAGES	SPAM	NOT SPAM	FALSE POSITIVE ALARM
Humanities	31,800	80%	12%	8%
Physical Sciences	26,512	74%	15%	11%
Engineering	20,280	78%	13%	9%

4.1.6 Challenge/Response Analysis

Challenge/Response analysis [24] is an intuitive method which carries out the sender's verification prior to the email delivery to the marked recipient. The sender of the message is sent a challenge e-mail in order to generate response to assess authenticity of the original message.

Table 6. Results for Challenge/Response Analysis

DATASET	TOTAL MESSAGES	SPAM	NOT SPAM	FALSE POSITIVE ALARM
Humanities	31,800	80%	12%	8%
Physical Sciences	26,512	74%	15%	11%
Engineering	20,280	78%	13%	9%

4.2 Spam Definition Database

The second most important component of the proposed anti-spam solution is a spam definition database. This module consists of all valid spam definitions that are recorded by spam analysis engine and are shared by online communities. This list can be generated exclusive to the very nature of the organization as based on the organization's perspective of spam. This database will include three fundamental lists:

a) White Lists — List of identities, usually of people working with the organization, and is issued by the official organization body

b) Policy Engine-based Exception Lists — These are generated based on specific message attributes that allow direct forwarding of the message to the network and then to the users. This type of mechanism is there to allow specific nature of messages to reach the server and the final recipient without any delay to expedite communication a specific nature related to the working of the organization. Policies can be generated for different tiers of an organization, also providing a dynamic link between the whitelists generated by different departments of an organization.

c) Black Lists: This list is generated for confirmed spam entries and entries that are considered incompatible with the exception policies of the organization.

4.2.1 Harvester Firewall

As spam analysis tool is mandatory for an effective anti-spam solution, it is ineffective without a strong firewall. An organization anti-spam framework must have a robust firewall that should have the capability to protect network and email databases from malicious attacks intended specifically for address harvesting. Spammers are often quite interested in harvesting e-mail addresses from an organization e-mail domain by checking responses from the server by sending fabricated messages to likely e-mail addresses present on that server. Based on rejected email returns, and applying the process of elimination, spammers manage to develop a directory including valid email addresses through successful hits. These addresses are then used to send spam emails and this directory is also forwarded to other spammers for selling purposes. An efficient firewall to prevent harvesting of emails is therefore quite necessary to implement an anti-spam solution.

Protection against Directory Harvest attacks [25] can be expedited if the firewall can handle and prevent address re-writing to obscure internal domains, Denial-of-Service (DoS) attacks, and open relay hijacking.

Results showed that almost half of the spam was blocked before analysis when a harvester firewall is used. This shows the magnitude of the harvesting attempts made by spammers. These results were recorded in parallel with the abovementioned results and all the emails were routed in parallel to the harvester firewall.

Table 7. Spam blocked by Harvester Firewall

DATASET	TOTAL MESSAGES	TOTAL BLOCKED
Humanities	31,800	33%
Physical Sciences	26,512	41%
Engineering	20,280	39%

4.3 Proactive Spam Blocking

Fourth component of the proposed anti-spam solution is proactive blocking of spam which can be quite complex when dealing with large number of incoming emails but it is one of the most effective anti-spam defenses that can help blocking spam at the Internet gateway. This technique will allow engaging spam prior to using the anti-spam analysis tool, thus reducing the overhead of further processing. Proactive blocking is proposed in this solution as it overwhelmingly improves spam processing/message throughput.

Proactive blocking in this solution was carried out using the SMTP relay [26], whose performance can be enhanced when used in conjunction with analysis tool.

Proactive blocking techniques include RDNS (reverse DNS lookup), RBLs (real-time black hole lists), and using rules from blacklists generated by firewall and analysis tool.

4.4 Identity-Based Spam Filter

Spam crisis is basically a crisis of fake identities and it exists mainly online because on the internet there are numerous ways to hide or fake one's identity. The SMTP protocol and the Internet allow anonymity which makes the law an underdog to the spammers as without identity there is no case. Therefore the legislative measures become useless as it is nearly impossible to find the spammer who desires to remain anonymous.

Only way around is authentication of senders identity versus a database of proven or valid identities. If email accounts can be linked with valid identities like social security numbers and identification tags, legal action can become more pronounced. Still in at an organizational level, a low level identity database can be maintained using assistance from international unions and through collaborations among commercial and private communities to dynamically maintain the identity/email-account database. By maintaining a known or trusted e-mail list from partners, customers, and other employees will improve spam analysis by reducing the load on the email server and the spam analysis tool.

Few commercial tools are currently available for identity validation/authentication and more work is being carried out to develop similar tools as a partial solution to problems generated by spam. The premise of this work is to develop solutions to validate identity using attributes of an email document, like digital signatures, encryption and certificates. Some key techniques that can be used for identity based filtering are [27]:

S/MIME –this is an Internet standard for e-mail encryption and it provides strong authentication using digital signatures. S/MIME is a feasible server based solution and it is extensively used in government sector, business, finance, and healthcare to develop secure networks and identities.

- TLS-based messages –TLS is a technique that allows direct communication between email servers after a security handshake is established for an encrypted channel.
- Directory integration –this technique verifies the address of the recipient before sending the email on to the network.
- Outbound e-mail recipient caching –This techniques uses information from outbound traffic to validate correspondents involved in an electronic communication.
- Recurrent harvesting of digital certificates from incoming emails to register and validate users for future inbound emails.
- Following actions were carried out by the spam analysis engine after analysis of the message content:
- White List: All email is tagged and delivered to respective mail accounts
- Black List: All email is tagged as spam and sent to spam storage.
- False Positive: Any email that is rejected by only 2 or lesser of the six analysis techniques and is not spam is marked as false positive.
- Spam: All emails tagged as spam by at least 3 of the 6 analysis techniques.
- Update Spam Database: New spam related attributes that are identified during analysis of incoming emails are stored at spam definition database.

Table 7 shows the overall performance of the proposed anti-spam solution. These results are based on a study carried out in three different departments in an educational institution over a span of one year. False alarm rate was reduced to less than 2 percent when the analysis techniques were applied sequentially.

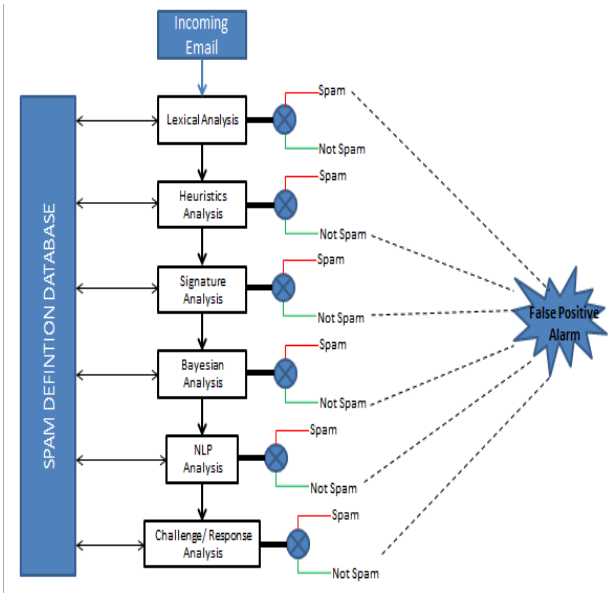

Fig3 : One of the rules for detecting false positive emails marked as spam

Table 8.Overall performance of proposed anti-spam solution

DATASET	TOTAL MESSAGES	SPAM	NOT SPAM	FALSE POSITIVE ALARM
Humanities	31,800	76%	23%	1%
Physical Sciences	26,512	74%	24%	2%
Engineering	20,280	79%	18%	1%

The aim here was to devise an anti-spam solution which can be adapted by any generic application belonging to industry, healthcare, education and government sector. As, the laws and legislations are quite explicit in defining spam and spamming activities, the measures to curb such activities can be carried out using the proposed anti-spam solution.

5. CONCLUSIONS

Spam is an infiltration of a user's privacy and afflicts financial penalties by occupying bandwidth and online resources. This problem can aggravate in enterprise based entities and organizations that do not have the luxury to make a mistake of marking a valid message as invalid. There is a dire need to develop a robust anti-spam solution which affectively and proactively detects spam and takes appropriate action without affecting the work process of an organization.

IT managers in these organizations are faced with the challenge of false positives from the valid emails which are business critical but can be attributed as spam due to the high volume of emails and spam definitions. Moreover, in limited time it is humanly impossible to detect, analyze and respond at the same time. And they must do this with limited human and financial resources. Secondly, numerous laws are being implemented worldwide to stop the influx of spam. These laws cannot become effective until suitable measures are taken by all stakeholders and spam is countered at every level. Moreover, such laws are difficult to apply as anonymity is the prime identity used by spammers. The solution to this problem is to develop an anti-spam model, which is generic, automated, and is applicable to small and large enterprises, and it not only helps in preventing spam inflow but also assists in stopping outflow of spam, that would have made the enterprise accountable to an anti-spam law. The proposed solution in this study is aimed at doing the same job but with practical means. The solution was tested in an educational organization and results acquired were quite promising. Results showed that spam analysis engine performance enhances when all the analysis techniques are used in combination sequentially. This greatly reduces the probability of false alarm and spam is efficiently discarded.

The five components proposed in this solution have shown to optimize spam blocking while minimizing false positives. The strategic placement of this solution at the network gateway will provide numerous advantages over standalone spam filters. These components can be used to develop a working solution which can then be modified according to the nature of organization and probable threat level of spam. A thorough testing of this framework is still pending, and it is currently being tested in an educational institution. Initial results were quite good, but a final deduction will be carried out after detailed tests and assessment of this solution in different real-time scenarios is completed.

6. REFERENCES

[1] Partridge, C., "The Technical Development of Internet Email," in Annals of the History of Computing, IEEE , vol.30, no.2, pp.3-29, April-June 2008

[2] Lei Jin; Takabi, H.; Joshi, J.B.D., "Security and Privacy Risks of Using E-mail Address as an Identity," in Social Computing (SocialCom), 2010 IEEE Second International Conference on , vol., no., pp.906-913, 20-22 Aug. 2010

[3] Aladdin Knowledge Systems, Anti-spam white paper, [Online]. Available: www.csisoft.com/security/aladdin/esafe_antispam_white paper.pdf [Retrieved 10-Aug- 2015]

[4] F. Smadja, H. Tumblin, "Automatic spam detection as atext classification task", in: Proc. of Workshop onOperational Text Classification Systems, 2002

[5] S. Hinde, 'Spam, scams, chains, hoaxes and other junk mail', Computers & Security, vol. 21, no. 7, pp. 592-606, 2002.

[6] M. Butler, 'Spam — the meat of the problem', Computer Law & Security Review, vol. 19, no. 5, pp. 388-391, 2003.

[7] S. Ghiam, 'A Survey on Web Spam Detection Methods: Taxonomy', International Journal of Network Security & Its Applications, vol. 4, no. 5, pp. 119-134, 2012.

[8] Acma.gov.au, 'Australian eMarketing Code of Practice | ACMA', 2015. [Online]. Available: http://www.acma.gov.au/Industry/Marketers/Anti-Spam/Ensuring-you-dont-spam/australian-emarketing-code-of-practice-ensuring-you-dont-spam-i-acma. [Accessed: 02- Aug- 2015].

[9] Lois-laws.justice.gc.ca, 'An Act to promote the efficiency and adaptability…' [Online] Available: http://lois-laws.justice.gc.ca/eng/AnnualStatutes/2010_23/FullText.html. [Accessed: 18- June- 2015].

[10] European Commission e-Privacy and Electronic Communications Directive. [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:02002L0058- 20091219: EN: NOT. [Accessed: 10-July-2015].

[11] Law.cornell.edu, 'U.S. State Anti-Spam Laws: Introduction and Broader Framework | LII / Legal Information Institute', 2015. [Online]. Available: http://www.law.cornell.edu/wex/inbox/state_anti-spam_laws. [Accessed: 24- June- 2015].

[12] Dia.govt.nz, 'Anti-Spam - dia.govt.nz', 2015. [Online]. Available: http://www.dia.govt.nz/services-anti-spam-index. [Accessed: 11- June- 2015].

[13] Mofo.com, 'Japanese New Anti-Spam Law | Resources | Morrison Foerster', 2015. [Online]. Available: http://www.mofo.com/resources/publications/2008/07/japanese-new-anti_spam-law. [Accessed: 01- July- 2015].

[14] Ida.gov.sg, 'Spam Control Framework - Policies and Regulations - Infocomm Development Authority of Singapore', 2015. [Online]. Available: https://www.ida.gov.sg/Policies-and-Regulations/Acts-and-Regulations/Spam-Control-Framework. [Accessed: 19- July- 2015].

[15] Oecd-ilibrary.org, 'OECD Anti-Spam Toolkit of Recommended Policies and Measures - Books - OECD iLibrary', 2015. [Online]. Available: http://www.oecd-ilibrary.org/science-and-technology/oecd-anti-spam-toolkit-of-recommended-policies-and-measures_9789264027176-en. [Accessed: 02-Aug-2015].

[16] Apec.org, 'APEC Principles for Action against Spam - Asia-Pacific Economic Cooperation', 2015. [Online]. Available: http://www.apec.org/Meeting-Papers/Ministerial-Statements/Telecommunications-and-Information/2005_tel/annex_e.aspx. [Accessed: 29- Sep-2015].

[17] Itu.int, 'ITU Activities on Countering Spam', 2015. [Online]. Available: http://www.itu.int/osg/spu/spam/intcoop.html. [Accessed: 18-Aug- 2015].

[18] Pages.au.int, 'Cyber Security | African Union', 2015. [Online]. Available: http://pages.au.int/infosoc/cybersecurity?q=infosoc/cybersecurity. [Accessed: 09- Aug- 2015].

[19] HailongHou; Chen, Yan; Beyah, R.; Yan-Qing Zhang, "Filtering Spam by Using Factors Hyperbolic Tree," in Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE , vol., no., pp.1-5, Nov. 30 2008-Dec. 4 2008

[20] Chi-Yuan Yeh; Chih-Hung Wu; Shing-Hwang Doong, "Effective spam classification based on meta-heuristics," in Systems, Man and Cybernetics, 2005 IEEE International Conference on , vol.4, no., pp.3872-3877 Vol. 4, 10-12 Oct. 2005

[21] Yan, J.; Pook Leong Cho, "Enhancing Collaborative Spam Detection with Bloom Filters," in Computer Security Applications Conference, 2006. ACSAC '06. 22nd Annual , vol., no., pp.414-428, Dec. 2006

[22] Issac, B.; Jap, W.J.; Sutanto, J.H., "Improved Bayesian Anti-Spam Filter Implementation and Analysis on Independent Spam Corpuses," in Computer Engineering and Technology, 2009. ICCET '09. International Conference on , vol.2, no., pp.326-330, 22-24 Jan. 2009.

[23] Kandasamy, K.; Koroth, P., "An integrated approach to spam classification on Twitter using URL analysis, natural language processing and machine learning techniques," in Electrical, Electronics and Computer Science (SCEECS), 2014 IEEE Students' Conference on , vol., no., pp.1-5, 1-2 March 2014

[24] Peng Li; Miao Liu; Fan Zhang; Tang Chunming, "An Improved Anti-Spam Method TOMO Based on Challenge-Response Technology," in Intelligent Ubiquitous Computing and Education, 2009 International Symposium on , vol., no., pp.113-116, 15-16 May 2009

[25] Suman Das; Singh, R.; Joshi, R.C.; Toshiwal, D., "Reducing the Effect of Distributed Directory Harvest Attack and Load of Mail Server," in Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on , vol., no., pp.1-6, 8-10 Dec. 2008

[26] Sandford, P.J.; Sandford, J.M.; Parish, D.J., "Analysis of SMTP Connection Characteristics for Detecting Spam Relays," in Computing in the Global Information Technology, 2006. ICCGI '06. International Multi-Conference on , vol., no., pp.68-68, Aug. 2006

[27] Wenxuan Shi; MaoqiangXie; Yalou Huang, "Collaborative spam filtering technique based On MIME fingerprints," in Intelligent Control and Automation (WCICA), 2011 9th World Congress on , vol., no., pp.225-230, 21-25 June 2011