I.J. Information Technology and Computer Science, **2016, 01, 11-21**
Published Online January 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2016.01.02

Study on the Effectiveness of Spam Detection Technologies

Muhammad Iqbal[1]
[1]School of Information Sciences & Technology, Southwest Jiaotong University, Chengdu, PR China
E-mail: muhammadiqbal@yahoo.com

Malik Muneeb Abid[2], **Mushtaq Ahmad**[3] **and Faisal Khurshid**[4]
[2]School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, PR China
E-mail: malik.muneeb.abid@hotmail.com
[3]School of Information Sciences & Technology, Southwest Jiaotong University, Chengdu, PR China
E-mail: mushtaq.ahmad91@gmail.com
[4]School of Information Sciences & Technology, Southwest Jiaotong University, Chengdu, PR China
E-mail: faisalnit@gmail.com

***Abstract*—**Nowadays, spam has become serious issue for computer security, because it becomes a main source for disseminating threats, including viruses, worms and phishing attacks. Currently, a large volume of received emails are spam. Different approaches to combating these unwanted messages, including challenge response model, whitelisting, blacklisting, email signatures and different machine learning methods, are in place to deal with this issue. These solutions are available for end users but due to dynamic nature of Web, there is no 100% secure systems around the world which can handle this problem. In most of the cases spam detectors use machine learning techniques to filter web traffic. This work focuses on systematically analyzing the strength and weakness of current technologies for spam detection and taxonomy of known approaches is introduced.

***Index Terms*—**Spam Detection Technologies, Machine Learning, Whitelists and Blacklists Signatures, Spam score.

I. Introduction

The definition of spam is not straightforward as this phenomenon is available in different forms and on different medias. One of the most widely recognized form of spam is email spam, the term is applied to similar abuses in other media: Internet, Cellular Networks and VoIP platforms. Fig. 1 describe further classification of this phenomenon. A more general definition for spam is described by Chan et al. [4], that a spam is an undesirable message sent to a recipient who has not requested it.

Today, most widely recognized form of spam is email spam. According to the Message Anti-Abuse Working Group (MAAWG) report [31], between 88–91% spam email messages sent since January 2012 to June 2014. The persistent presence of unwanted internet traffic is a vigilance signal to the industry and researchers to remain open-eyed against this issue.

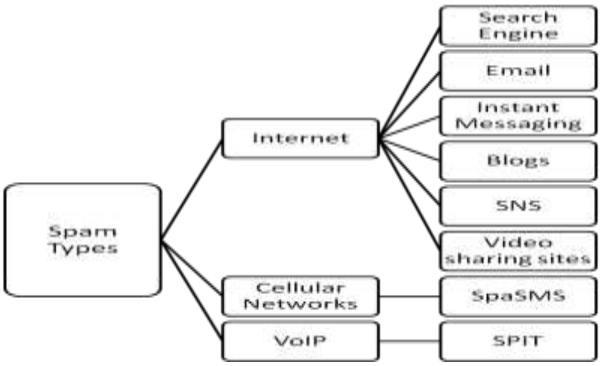

Fig.1. Different Types of Spam

Although most of the recent computer networks are being properly designed and segmented accordingly, however, still there is a risk of targeted spam attacks. In order to protect from un-wanted traffic like spam, it is significant to know which endpoint and network based cyber security controls are available [9, 10].

Currently there are two main independent battle fronts have been opened to fight against web spam i.e., legal and the technological [32]. The legal front has encountered some difficulties due to the international character of Internet. A comprehensive anti-spam legislation could also help to deal with this problem. On technological front several measures have been devised and put into practice by companies.

A. Taxonomy of Web Spam

Due to financial aspects of spamming, presently there are many types of electronic spam, including spam by instant messaging (spim or SpaSMS), spam by internet telephony (spit), spam by mobile phone, by fax, Web spam and etc. The Web spam is not absent from this list, but as the request response paradigm of the HTTP protocol makes it impossible for spammers to actually "send" pages directly to the users, spammers try to deceive search engines and as a result break the trust that search engines establish with their users. All deceptive

actions which try to increase the ranking of a page in search engines are generally referred to as Web spam or spamdexing (combination of "spam" and "index") [1]. In another definition of spam, the authors [3] describe spam as "Web spamming refers to those unethical actions of spammers who intended to mislead search engines and give some pages higher ranking than they deserve". There are four famous Web Spam techniques (See Fig. 2) which are currently challenge for search engines i.e. Content Spam, Link Spam, Cloaking & Redirection and Click Spam [21].

In content spamming, the spammers play with textual features to subvert the ranking of search engine. Link spamming refers to any web spam method that tries to improve the link-based score of an intended web page by creating lots of hyperlinks pointing to it.

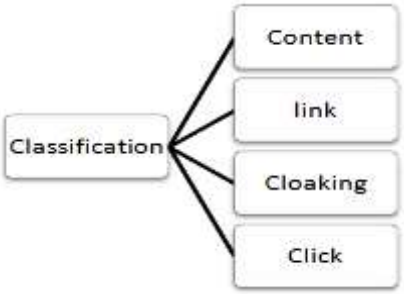

Fig.2. Classification of web spam

Cloaking refers to spam technique to serve a page to the search engine spider that is different from that seen by end users. In click spamming, a method is used to submit the queries to search engines that retrieve target result pages and then to "click" on these pages in order to achieve spammers objectives.

B. Spam in Cellular Networks

The use of smart phones has seen an adoption rate faster than any other technology in human history [19]: as in 2012, the number of cellphones subscribers had almost exceeded 6 billion users. This domain also witnessed the fast rate of innovations. Companies like Google, Microsoft, Apple, Samsung and Xiomi have all provided APIs for enabling open application development on the cellphones. These cellphones are now empowered with different applications, including location based services. Cellular networks have been playing very important role for information interaction, resource sharing, and social communications. Significantly, they have become a vital platform for the provision of various applications and services, such as mobile Email, Short Message Service (SMS), mobile commerce, multimedia communications and mobile social networking etc. These wide ranges of applications on cellular networks bring us a great convenience and become more and more indispensable in our modern life. Currently, billions of cellular devices are being connected through Internet and access its services and applications, which satisfy people's needs and ease their life [5]. According to Cisco Internet Business Solutions Group (IBSG) study [20], Cisco predicts that there will be almost 25 billion devices connected to the Internet by the end of 2015. It is significant to note that these estimate figures do not take into account rapid advances in Internet or device technology; the numbers presented are based on what is known to be true today.

However, the cellular networks are also becoming a main source to transmit a lot of unwanted contents. The unwanted content could be malware, virus, spam, intrusion, and unsolicited commercial advertisement. Thus, it could greatly bother mobile users and at the same time could infect their devices. These unwanted contents only benefits its source, but burdens both end users and network service providers by adding extra load into the network, which greatly leads to network traffic congestion. It consumes network and computing resources in a way that does not benefit its receivers, thus it should be ascertained, filtered or discarded during transmission from its source to destinations.

Since 2001, China has become one of the largest and fast growing cellular telecommunications market. The statistics shows that till February 2015, about 1.29 billion mobile phone users had been registered in China with different telecom operators and on an average each mobile user received 10.7 SMS spam messages on average per week [6].

C. Problem Statement

Distinguishing spam and ham traffic is a complex and constantly challenging task. Developing a model to classify the broad range of spam types is tedious job; this task is made near impossible with the realization that spam types are constantly moving and evolving. Spammers use different unethical techniques for altering their messages to avoid detection system and adding a further hindrance to accurate detection system [22]. In order to deal with this massive problem, pretty sophisticated technologies are being developed by researchers and that are currently in use of IT companies. A maximum number of these technologies operate in the background without the system administrators even knowing what is going on. Fighting against spam must be effective and easy especially for small IT companies where there are few people with limited resources and they don't have the time to specialize in one aspect of technology.

In order to protect from objectionable (unwanted) contents, different content filtering techniques are being used by companies to screen and exclude from access or availability to end users. There is a acute need of study of currently employed anti-spam technologies which have been used to fight spam using automated solutions. Moreover, a study which will help system administrators and researchers to expand their knowledge base in this domain.

D. Proposed Solution

Keeping in view the negative impact and scale of problem, this paper conducts a study of different anti-spam filtering methods. This paper provides the knowledge of good spam classifier to protect companies

and individuals from this problem. We have carried out comparative evaluations of well known employed anti-spam algorithms to judge their algorithm properties. Moreover, this paper discusses the causes/sources of this phenomena and review of Machine Learning (ML) and Non-Machine Learning (NML) methods. A comparison of different approaches is also presented in tabular form for quick understanding for end users. An experimental work is done on publically available dataset (webspam-uk2007) to judge the performance of well known anti-spam machine learning algorithms (J48 and NB) by taking True Positive (TP) and False Positive (FP) metrics.

E. Organization of paper

This paper is organized as follows. In the rest of the paper, we first discuss the spam implication and associated statistics and their problems in Section 2. Section 3 describes the performance measurement criteria. Section 4 reports the spam detection methods. In section 5, we compare the advantages and disadvantages of different spam filtering techniques. Experimental work and results have been discussed in Section 6 and in Section 7 we conclude the paper and define some future directions..

II. Overview of the Problem

The first demonstrations of sending spam messages towards users groups are dated from the early seventies, when a revolution begun in electronic communication due to World Wide Web (WWW) presence as communication platform. In the present age, we are witnessed of great transformation of WWW lasting for almost five decades, which produced business policy paradigm shift, and affecting mainly commercial background and searching for high financial profits. Currently majority of business companies are pretty much depending on this platform to reach their customers worldwide.

Spamming becomes an attractive phenomenon for spammers due to below five main reasons [21].

1. Financial benefits to be earn from search engines
2. It sabotages the trust of end user in a search engines
3. Spam websites serve as means of disseminating malware and adult content dissemination and also being used for fishing attacks.
4. Search engine may spend large amount of computational and storage resources on spam pages.
5. Declining employee productivity is the overwhelming by-product of spam.

Table 1. depicts the core reasons of using Internet platform for their businesses:

Table 1. Reasons to adapt Internet platform for business purpose

Rank	Adopting Drivers
1	Scalability
2	Direct and indirect advertising
3	Low cost communication
4	Easy access to potential customers
5	Better Management
6	Company value Enhancement

The proper use of Internet technologies, like teleconferencing, email systems and different automation systems appear to have to improve organizational performance by lowering costs, increasing efficiency, differentiating products and services, or creating broader markets. The effective use of Internet technologies can bring a sound impact on organizational structure and its functions.

In these days, the use of ecommerce technology is flourishing with fast pace in developing and developed countries. If we just take a case of Peoples Republic of China (PRC) then since 1993, the people of PRC are also gaining benefits from this boom, when the foreign businesses in China started to use EDI to simplify trading processes [13]. Latter on Chinese businesses also began to adopt this new technology. According to one study [14], in 2013, the growth of business applications increased with fast pace. The figures of online shoppers in PRC reached 302 millions. In 2013, the online procurement and online sales of Chinese enterprises were reached 23.5% and 26.8% respectively, and the proportion of the enterprises that launched Internet-based marketing promotion campaigns was 20.9%.

The above statistics of internet usage shows the business trends of just one example of single country (PRC), where economy is boosting due to prominent role of Internet platform, otherwise the web made e-commerce an easy and cheaper way of doing business and enabled more diverse business activities. It is difficult to list all possible applications of Internet which are being exercised by the users, but few are: Education, Business, Communication, Leisure, Medicine etc.

At the same time, this platform appears with new challenges like spam to individuals and to companies. Different very worrying reports[15,16,17] revealed from companies like Microsoft, Kaspersky, Symantec, indicates an increasing percentage of spam cases in our mail, dangerous revival of phishing attacks and viruses' transfer are the most zealous propagators. The total number of electronic messages sent on a world scale, reaches the threshold of around 200 billion per a day and the current share of spam in this amount is estimated to 90% [18].

In most of the cases the majority of recipients reply to annoying contents, that involves expenses of billions of

dollars on operation and as a consequence, reduction in efficiency of work. According to the networking giant CISCO [18], the USA has a solid lead in this top spammers' list. Fig.3 shows that over 17% of all spam traffic comes from America. Turkey is a runner-up contest with around 9% of worldwide spam producer. Russia got third rank with 8% spam producing country. The rest of the top spamming countries are far behind:

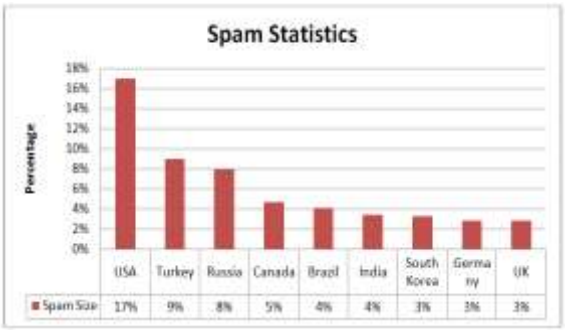

Fig.3. Spam relaying countries data [18]

According to CISCO [18] the global spam volume is falling according to data collected by Cisco Threat Research Analysis and Communications report for the year 2013, although trends vary by country.

A major source of spam comes from advertisements on Internet and these adds mostly belongs to the a wide range of pharmaceutical offerings like weight loss, height increase, sexual aids, pain relief, sleeping aids and sexual health. Additionally finance, software's and music domains are also favorites for spammers. The key principle of marketing opinion is based here on luring offers of attractive products or services at cheap prices.

III. Performance Measurement Criteria

It is very unfortunate that the spam is keep growing each year with a fast pace. According to different studies [15, 16, 17, 18], the volume of spam worldwide has been increasing in all internet traffic, and is continuously increasing with each passing year. Different techniques are being implemented by spam detecting algorithms to eliminate spam and every detector is achieving different performance levels. The classification task to distinguish spam and ham is complex and constantly changing. Due to this nature of the problem; unfortunately, till to-date no exact solution has been explored by researchers to eliminate spam traffic [25]. However, majority of spam detectors have two main attributes in common that determine their overall efficiency:

- Number of spam messages detected
- Number of ham (legitimate) messages falsely reported as spam.

The first measure is Detection Rate (DR), and the second is known as the False Positive Rate (FPR). DR is the total number of messages declared as spam by the spam detector compared to the number of all spam messages in email traffic over a fixed period of time. For example, if an organization or an individual receives 100,000 spam messages over the period of one week, and 98,000 were eliminated /blocked by the spam detector, the detection rate is 98%. Nowadays, a spam detector considers effective if it has 95% or better DR value.

$$DR = \frac{Sc}{Ts} \quad (1)$$

Where
Sc = Number of Spam correctly Specified
Ts = Total number of all spam messages over a certain time period.

It is important to note that DR is not the only criteria to judge the performance of anti spam solutions. Users should also look the false positive rate of spam detector. In false positive rate (FPR) the message is mistakenly tagged as spam and ultimately brings a serious problem to end users.

$$FPR = \frac{Fp}{Th} \quad (2)$$

Where
Fp = Number of false positive
Th = Total number of all ham messages over a certain time period.

FPR is very important factor, which should be taking into account to measure its performance. For example, frequently tagging ham message as spam could result in big problem for company users. A small number of spam messages a day is a small price to be paid, but missing an important email can result in serious consequences to company or user.

An effective spam detector should demonstrate false positive rate (0.001%) or less, i.e. one false positive per every 100,000 messages. Unfortunately, till to-date, no anti-spam software vendor claims that their false positive rate is zero.

IV. Spam Detection Methods

In order to detect spam traffic, different detection methods are being developed by researchers. In this section we will present the widely used methods. These detection methods can be categorized in two approaches (1) machine learning (ML) rules and (2) those not based on ML (see Fig. 4 (a)). Machine Learning and Non-Machine Learning approaches are further divided into sub categories (see Fig. 4 (b) and 4(c)).

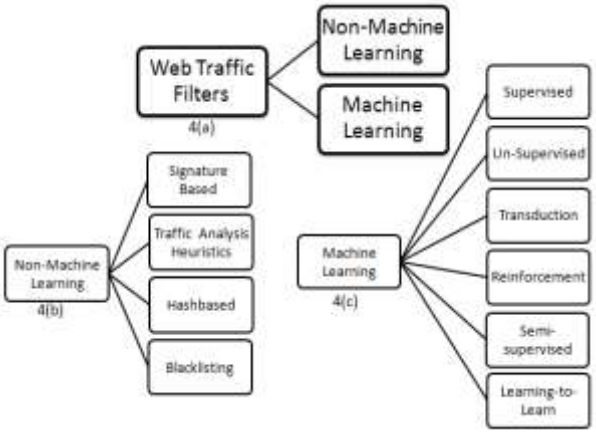

Fig.4. Spam detection methods categories

A. Challenge Response Authentication

In Challenge-Response-Authentication (CRA) session one entity send a challenge to another entity through its family protocols. In response, the second entity must respond with the appropriate answer to be authenticated. Password authentication is a simple example of this method. In this fashion the challenge is from a server asking the client for a password to authenticate the client's identity so that the client can be served. The second entity should respond with the exact answer to be authenticated.

CRA is dependent on two entities (1) a secret value, and (2) a variable challenge value. One of the strength of CRA is the submitted authentication value is always different each time and dependent on a challenge value, so it is more difficult for an attacker to replay a previous authentication value [27].

One common method for fighting spam is frequently used to prevent spam sent via contact forms on the web. When end users fill out a form, the CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) [7] test that he/she must successfully enter before hitting. In CAPTCHA test (see Fig. 5), a computer program generates and grade test that most humans can pass but computers generated codes cannot pass.

Fig.5. CAPTCHA test example

Algorithm 1: Challenge Response Authentication Algorithm

Step 1:Client → Request for Service
Step 2: Server → Passes the stored, challenge(Value) to the client
Step 3: Client → Computes response = Answer(Value), passes it to server
Step 4: Server → Checks if ((Answer(Value) = Challenge(Value) = = Response), which will mean successful authentication, and if so proceeds with

A good number of websites are already deployed CAPTCHA tests during registration and initial login from end users. Nowadays, several DDoS victims use CAPTCHA technology to protect themselves against application-layer DDoS attacks such as HTTP flood attacks [8]. Every day millions of CAPTCHAs are solved by humans around the world. A standard algorithm for challenge response authentication process is described in below algorithm1.

Due to three important security features of CRA i.e. authentication, prevention from replay attacks and secrecy; nowadays majority of smart card systems use challenge-response authentication approach. These systems require at least two things for authentication and entry from users: the smart card and the user's password (PIN CODE).

Another example of CAPTCHA is a form of reverse-Turing test for the system to determine if the client is a human or not. This is used to prevent spam and auto-registration of new accounts for a website or email. The use of biometric systems [9] is another form of challenge-response authentication. In cryptography, zero-knowledge password proof and key agreement systems such as secure remote password, CRAM-MD5 and secure shell's challenge-response system based on RSA are considered to be use of very advanced challenge-response algorithms.

Below are major applications of challenge response authentication systems

- Protecting Website Registration
- Preventing Comment Spam in Blogs
- Online Polls
- Preventing Dictionary Attacks
- Protecting Email Addresses From Scrapers

B. White-Lists and Black-Lists

White-lists and black-lists are considered to be the initial techniques deployed to stop spam. It works on a principle that the words or patterns which define a message as 'ham' (legitimate email) are white-listed and those which define a message as spam are black-listed. The body content and header of each email message is analyzed against these lists and the message is sent or

blocked accordingly by the algorithm. Thiago et al. [24] reported 4 steps: i) tokenization, ii) lemmatization, iii) stop-word removal and iv) representation, for extraction of data from message as a data per-processing step. Dinha et al. [26] proposed multiple features from header and body of spam message.

This concept is implemented through content, IP, MAC etc to block or allow data traffic in networks. Currently white lists and black lists are considered to be least effective methods for detecting spam. In many cases, these lists hurt innocent people and prevent critical business e-mail from being delivered. Algorithm 2 describes the general implementation of various lists for incoming internet traffic.

Algorithm 2: Lookup order of various lists during processing of the incoming messages

Step 1: All lists are checked against incomming message:{ Where List = white/black/dynamic Lists etc }
Step 2: Compute threashold value of message
Step 3: Check, if Threashold Value (T_h) >= Desired parameter Value (D_v) Step 3.1: Declare Spam and Move message to spam label Step 3.2: else Process message
Step 4: Go to step 1.

For example, if system is configured with black-list and message contain word 'sex' might also block the word 'Middlesex'. Another problem is that spammers can change words to fool the blacklists, for example, 'Vi@gra' is used in place of 'Viagra'. Hence, blacklisting and white-listing systems are largely ineffective, as they could be spoofed and changed by spammers[10].

Another example of weaknesses of this technology can be seen in DNSBLs(DNS Blacklists) where the email servers keep the list of IP addresses, published by a third party, and spam filter can only stop spam if the source address has been available in the list. Domain based DNSBLs are also known as Right Hand Side (RHS) black lists because these lists look at right hand side of @ sign. For example a email came from xyz@abc.com ,then this kind of lists can only check at abc.com. If the domain name is part of blacklist then the email from this domain will be blocked, otherwise it will be received by recipient.

Some time legitimate email domains victims of these list. In many cases the smart spammers simply alter their sending DNS to fool the filter.

C. Pattern Detection

Pattern Detection technology is a combination of methods, like operation research, graph theory, data analysis, clustering and advanced mathematics for extracting meaning from large and complex data sets. Fig. 6 depicts exploded working of pattern detection methods can identify commonalities among spam messages because it is relying on a big database of spam messages collected all over the world to determine what institutes spam. The learning process to indentify the different techniques used by spammers makes it one of the most advanced spam fighting technologies in use today.

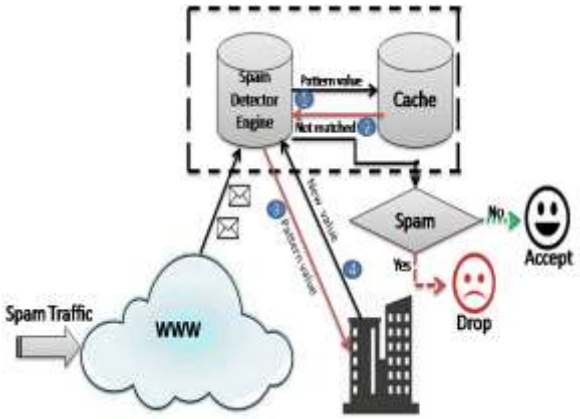

Fig.6. General Framework of spam detection through pattern detection

Algorithm 3 shows the methodology to implement rule based filtering approach.

Algorithm 3: Rule Based Filtering Algorithm

Step 1: Start
Step 2: Spam data arrives at spam detection System
Step 3: Spam detecting Engine sends Message Pattern Characteristics to its Local Cache Step 3.1: Local cache stores the spam patterns of all the recent attacks. Step 3.2: If a message pattern characteristic is found, Goto step 6
Step 4: If the matching pattern is not found in Cache Step 4.1: Message Pattern Characteristics is send to the remote Spam Detection Center.
Step 5: Detection Center classifies the message and sends reply to Anti Spam Engine
Step 6: Anti Spam Engine forwards message to the mail recipient if it is not spam else it will drop the mail.
Step 7: Algorithm stores the newly classified pattern in its local cache for future use
Step 8: END

Nowadays, this is one of the most prominent methods for fighting spam. In this technique, the rules are being framed by administrator that tells the spam filter about what to block. For example, when the words "Free preview" appears in a message, the spam filter knows to block that message because it violates a rule set by the administrator. A good number of Information Technology (IT) companies who are engaged in developing anti spam solutions provide pre-set rules for the people who use their products.

For the last years, anti-spam rule-based systems (RBS) has become popular in the filtering industry due to their

ability to successfully combine different classification techniques and the possibility of updating filters remotely. In this context, SpamAssassin is playing a vital role to develop this kind of filters. It is one of the hybrid filtering methods, uses content-based filter and real-time blacklists. It has been adopted by international companies (such as Symantec or McAfee) and small and medium enterprises (SMEs) [11].

Anti-spam RBS are combination of a decision threshold value plus a set of scored rules (See Fig. 7). Each rule holds a logical test (rule trigger) and a numeric (positive or negative) score. In this type of filters, when a rule matches the target message, its score is added to a global counter. After examining all rules, a message is classified as spam if its global counter value is greater than or equal to the configured threshold. Both (i) scored rules and (ii) the filter threshold are commonly stored in regular text files in order to facilitate their exchange between computers [11].

Algorithm 4: NB filtering

Training Phase Step 1: Create spam and ham sets by collecting many e-mails
Step 2: Retrieve individual tokens strings as feature words:
Step 3: Calculate the appearance time of the token and build the feature set $f = \{w_1, w_2, \ldots, w_n\}$.
Step 4: Generate hash tables for both ham and spam for the mapping relation of a feature word tokens.
Step 5: Compute the class-conditional probability P(w t |c i) for feature word w t
Classification Phase Step 1: Retrieve feature words from new message.
Step 2: Calculate the probability $P(c_{ham}\|d)$ of legitimate message and $P(c_{spam}\|d)$ of spam when it satisfies the extracted feature words d.
Step 3: Classify the incoming message based on the results. When the value of $P(c_{spam}\|d)$ is greater than $P(c_{ham}\|d)$ or the threshold value λ, this e-mail is tagged as spam.

D. Statistical Content Filtering

Numerous studies [33, 34, 35] have been done to obtain better spam detection results, however, the statistical approaches [36] use machine-learning algorithms to classify spam messages.

Bayesian network can be described as a graphical model for probabilistic relationships among a set of variables. Over the last decade, the Bayesian network has become a popular representation for encoding uncertain expert knowledge in expert systems.

Naïve Bayes filtering technique is widely used machine learning method. Formulation of NB is a combination of training and classification or testing and is described in Algorithm 4.

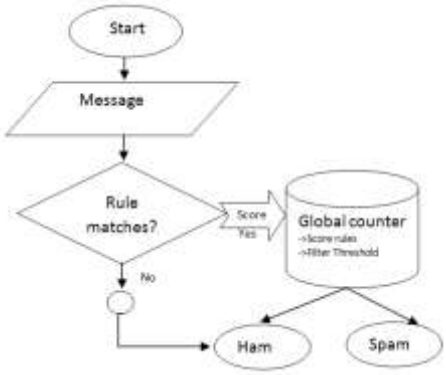

Fig.7. Flow chart of RBS working

Bayesian filtering is based on the principle that most events are dependent and that the probability of an event occurring in the future can be guessed from the previous occurrences of that event [12]. In Bayesian net(Bayesian filtering), it assign a score to different tendencies used by spammers. For example, a message with a high percentage of misspelled words sent from an Russian IP address that mentions Viagra (or spelling variations) has more tendencies used by spammers than a message regarding your annual sales forecast. Seeing that the first message fits a specific pattern, that message would be blocked if the score meets the threshold set by the administrator.

E. Database creation for filtering

In order to use Bayesian filtering method, the user needs to generate a database (see Fig. 8) with words and tokens (such as the $ sign, IP addresses and domains, and so on), collected from a sample of spam mail and legitimate mail.

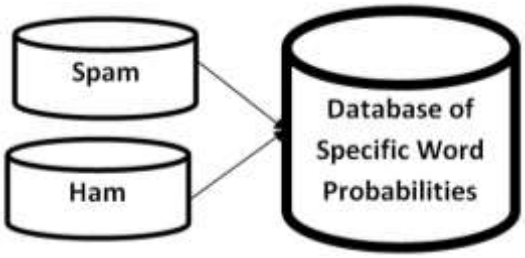

Fig.8. Database creation for filters

A probability score value is then assigned to each word or token; the probability is based on calculations that take into account how often that word occurs in spam as opposed to legitimate mail (ham). This process can be carried out by analyzing the end users outbound mail and by inspecting known spam: All the words and tokens in both pools of mail are analyzed to generate the probability that a particular word points to the mail being spam. This word probability score would be calculated as follows:

If the word “cheap” occurs in 500 out of 3,500 spam mails and in 5 out of 300 legitimate emails, for example,

then its spam probability would be 0.8955 (that is, [500/3500] divided by [5/300 + 500/35000]).

$$\Pr(\text{spamwords}) = \Pr(\text{word}(s\ ) \text{ spam}) / ((\Pr)(\text{words} \mid \text{ham}) + \Pr(\text{words} \mid \text{spam})) \quad (3)$$

Algorithm 5: RBF Method to catch spam

Step 1: Start
Step 2: Retrieve new message Step 2.1: Break Message in to Tokens Step 2.2: Calculate ProbScore(Tokens)
Step 3: Consult Ham and Spam Database
Step 4: If ProbScore(Tokens) > Threshold Value Step 4.1: Declare Spam ,Otherwise HAM
Step 5: END

The updating frequency mechanism of database file for Bayesian filter makes it secure to end users. This spam data file must hold large sample size of known spam and must be constantly updated with the latest spam by the anti-spam software. This will assure that the Bayesian filter is aware of the most recent spam tricks, and ultimately resulting in a high spam detection rate.

General Algorithm by using statistical content to filter spam traffic contains the following steps:

Following are some important reasons to choose Bayesian Filters to detect spam traffic:

- Bayesian filtering apply intelligent approach to filter data because it examines all aspects of a message, as opposed to keyword checking that classifies a mail as spam on the basis of a single word.
- A Bayesian filter is constantly self-adapting - By learning from new spam and new valid outbound mails, the Bayesian filter evolves and adapts to new spam techniques.
- The Bayesian method is multi-lingual and international

V. Comparison of Several Techniques for Spam Detection

Table 2 shows a summary of productivity and limitations of several spam filtering methods. Conducted studies indicate that machine learning techniques are more flexible than other methods. Pattern detection techniques and Statistical content filtering methods have achieved good performance scores and thus widely used in industry as compared to traditional (CRA and various listing) methods.

Table 2. Summary of several spam filtering techniques

Technique	Pros	Cons
CRA	Widely used protocol for unsecure channels	Weakness in authentication. Security and QoS issues ,especially in Wireless networks [28]
Black/ White Listings	Blacklisting/Whitelisting is a simplistic technique that is common	It can be easily penetrated into systems, ad do suffer with high rate of false positives [22]
RBS	Hybrid Model (content-based filter and real-time blacklists) Ability to successfully combine different classification techniques and the possibility of updating filters remotely[29]	Filtering speed is currently very limited[29]
Statistical Content Filtering	Hybrid Model (content-based filter and real-time blacklists) Ability to successfully combine different classification techniques and the possibility of updating filters remotely[29]	Filtering speed is currently very limited[29]

VI. Experiment and Results

In this section, we demonstrated the performance of two most widely and accepted supervised machine learning methods i.e. NB and J48 to detect spam pages. For performance evaluation, we have used WEBSPAM-UK2007 dataset [30]. Table 3. depicts the important information about dataset.

Table 3. The Properties of the WEBSPAM-UK host graph

Year	2007
Number of nodes(hosts)	114,529
Number of edges	1,836,441
Number of labeled hosts	6,479

The data set used in our experiment has 3851instances in which only 3% are spam and each instances has 142 attributes.

The hosts of this reference collection were originally labeled as “normal”, “borderline”, and “spam” by a group of volunteers. Each corpus host was labeled by at least two persons independently. Through data preprocessing phase, we have separated ham and spam pages for our experiment work.

The dataset contains content and link based features. Some of the important content features are “number of words in the page”, “number of words in the title”, “average word length”, “compression ratio”, “entropy”, “fraction of anchor text”, “fraction of visible text”, etc. These features are calculated for home page, page with

maximum PageRank, and an average value for all pages of every host. We have used 10-fold (see Fig. 9) embedded cross-validation during ensemble training and building.

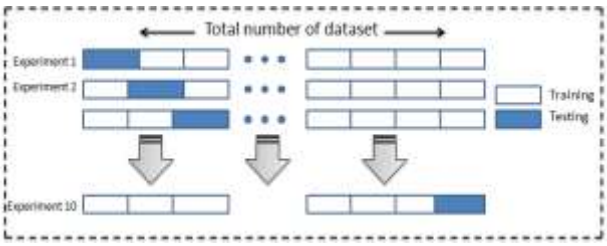

Fig.9. Example of 10-fold cross validation

We applied Naïve Bayes and J48 classifier on our dataset to judge algorithms performance by using True positive (TP) and False Positive (FP) scores in spam detection. For performance analysis we have used Weka toolkit. Table 4 shows the TP and FP scores for each algorithm with different feature values. The matter of fact is there is no universally best learning algorithm that map on every problem. In order to obtain promising scores [23], machine learning algorithms are mainly dependent on number of parameters like; size of dataset, feature selection and feature extraction techniques etc. In our experiment (see Fig. 10 and 11) J48 produced better TP and FP score comparable to NB, but with median number of features NB achieved good FP value(see Fig. 10). Although J48 performed well, but as we increase the number of features it is prone to decrease its TP score. We noticed that, high performance with minimum features was achieved with both j48 and NB algorithms.

Table 4. TP and FP scores

No. of Features	J48		NB	
	Measure			
	TP	FP	TP	FP
20	0.999	1	0.968	0.952
50	0.979	0.817	0.947	0.947
80	0.977	0.731	0.902	0.779
139	0.976	0.76	0.21	0.111

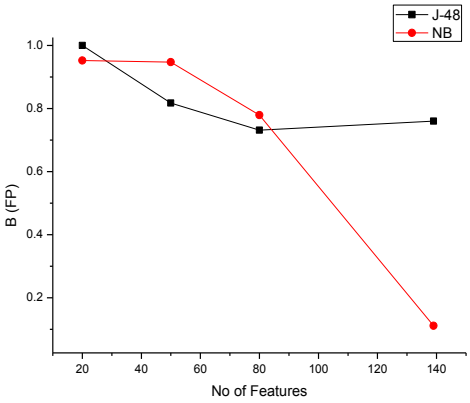

Fig.10. False Positive score comparison

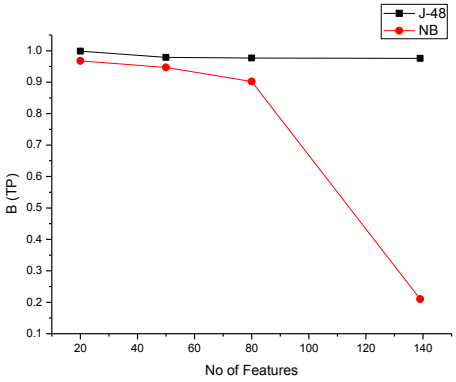

Fig.11. True positive score comparison

VII. Conclusions

This paper discusses the spam detection techniques which increase in effectiveness in the face of consolidating their power and robustness. The arsenal of new solutions to fight against these phenomena evolves as dynamically as the spam tricks, with the most important statistical analysis or machine learning. The existing architectures dividing the decisive factor into several points on the way of internet traffic and researchers are working on developing new intelligent methods to deal with this problem. The negative impact of spam brought the serious challenges to IT companies and end users and this situation compels involved companies to deepen social awareness, to outline the threats and draw public attention to the best practices of rules of proper prevention and defense.

We reviewed the different techniques to deal with spam problem based on traditional schemes (challenge response authentication and White/Black Listing) and different pattern recognition methods and statistical content filtering approaches proposed so far. We have precisely discussed the main approaches, on which these algorithms are based, as well as their weaknesses and strength for detecting spam. In order to convert perceptual level into practical, we have investigated two most widely used ML algorithms on webspam-uk2007 dataset. Even dataset is highly imbalanced, however, J48 achieve better TP score against NB.

There is no exact solution to curb spam problem, however the only real solution to this problem is to combine techniques, to create a probability score to assess whether an incoming traffic is spam. Nowadays, spam problem exists on almost all major mediums. Due to financial motivations, spammers always try to find new ways around detection methods. In order to protect networks and users from spam, the detection systems must be able to continuously learn and adapt new rules.

Acknowledgment

I would also like to thanks the other PhD Scholars of my school, Mr.Amjad Mehmood, and Mr.Mehtab Afzal for the assistance they provided to understand machine

learning. A very special thanks goes out to Dr. Zhu Yan, without whose motivation and encouragement, I confess that it would be difficult for me to move forward in my PhD Program.

REFERENCES

[1] D. Saraswathi, A. V. Kathiravan, R. Kavitha. Link Farm Detection using SVMLight Tool. In proceedings of ICCCI 2012 International Conference on computer Communication and Informatics (ICCCI -2012), Jan. 10 – 12, 2012, Coimbatore, INDIA.

[2] Eiron, N., Curley, K. S., and Tomlin, J. A. 2004. Ranking the web frontier. In Proceedings of the 13th international conference on World Wide Web. ACM Press, New York, NY, USA, 309–318.

[3] Z. Gyongyi, H. Garcia-Molina, Web Spam Taxonomy. In First International Workshop on Adversarial Information Retrieval on the Web, 2005.

[4] P. P.K. Chan,C.Yang,D. S. Yeung,W. W.Y. Ng, Spam filtering for short messages in adversarial environment, *Neurocomputing*, Vol. 155, 1 May 2015, pp. 167–176.

[5] L. Chen, Z. Yan, W. Zhang, R. Kantola, TruSMS: A trustworthy SMS spam control system based on trust management, Future Generation Computer Systems, Vol. 49, August 2015, pp. 77–93.

[6] China: mobile phone subscribers by month February 2015 [online], Available: http://www.statista.com/statistics/278204/china-mobile-users-by-month, June 11, 2015.

[7] L. Ying-Lien, H. Chih-Hsiang, Usability study of text-based CAPTCHAs, *Displays*, Vol.32, no. 2, pp. 81–86, April 2011.

[8] H. Beitollahi,G. Deconinck, Analyzing well-known countermeasures against distributed denial of service attacks, Computer Communications ,Vol. 35, no. 11, 15 June 2012,pp. 1312–1332.

[9] J.K. Dharavath, F. A. Talukdar, R. H. Laskar, Study on Biometric Authentication Systems, Challenges and Future Trends: A Review , International Conference on Computational Intelligence and Computing Research (ICCIC), 2013 IEEE Enathi.

[10] S. Heron, Technologies for spam detection, Network Security Vol. 2009, no. 1, January 2009, pp. 11–15.

[11] D. Ruano-Ordás,J. Fdez-Glez,F. Fdez-Riverola, J.R. Méndez, Effective scheduling strategies for boosting performance on rule-based spam filtering frameworks, Journal of Systems and Software ,Vol. 86,no. 12,, December 2013,pp. 3151–3161.

[12] Z. Qingshan , W. Shaobing, C. Ying, J. Xueming, The Research of Information Filtering Technology Based on Bayesian Network, Procedia Environmental Sciences , Vol. 11 (2011) , pp.545 – 551.

[13] X. Du, Internet adoption and usage in China, Proceedings of the 27th Annual Telecommunications Policy and Research Conference, Alexandria, VA (1999).

[14] CNNIC 33rd Statistical Report,[online], Available: http://www1.cnnic.cn/AU/MediaC/rdxw/hotnews/201401/t20140117_43849.htm, June 15,2015.

[15] Spam and Phishing Statistics Report Q1-2014,[online], Available: http://usa.kaspersky.com/internet-security-center/threats/spam-statistics-report-q1-2014#.VXHbzHKSwQU, June 16, 2015.

[16] 2015 Internet Security Threat Report Vol.20,[online], Available: https://www4.symantec.com/mktginfo/whitepaper/ISTR/21347932_GA-internet-security-threat-report-volume-20-2015-social_v2.pdf, June 18,2015.

[17] Microsoft Security Intelligence Report Vol. 18, [online], Available: http://www.microsoft.com/en-us/download/details.aspx?id=46928, May 14, 2015.

[18] Cisco 2014 Annual Security Report,[online], Available: http://www.cisco.com/web/offer/gist_ty2_asset/Cisco_2014_ASR.pdf May 4, 2015.

[19] N. Eagle, A. Pentland, Social serendipity: mobilizing social software, IEEE Pervas. Comput., Vol. 04-2 (2005), pp. 28–34

[20] D. Evans, The Internet of Things How the Next Evolution of the Internet Is Changing Everything, Cisco Internet Business Solutions Group (IBSG) April 2011[online],Available: http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf, 2 April 2015

[21] N. Spirin, J. Han, Survey on Web Spam Detection: Principles and Algorithms, *SIGKDD Explorations*, Vol. 13, no. 2, pp. 50-64, 2011

[22] J. Carpinter, R. Hunt, Tightening the net: A review of current and next generation spam filtering tools, *Computers & Security*, Vol. 25, no. 8, pp. 566–578, November 2006

[23] C. Laorden, X. Ugarte-Pedrero, I.Santos, B. Sanz , J. Nieves, P. G. Bringas, Study on the effectiveness of anomaly detection for spam filtering, Information Sciences Vol. 277, pp. 421–444, September 2014

[24] T.S. Guzella, W. M. Caminhas, A review of machine learning approaches to Spam filtering, *Expert Systems with Applications*, Vol.36, no. 7, pp. 10206–10222, September 2009

[25] M. A. Al-Kadhi, Assessment of the status of spam in the Kingdom of Saudi Arabia, *Journal of King Saud University Computer and Information Sciences*, Vol. 23, no. 2,pp. 45–58, July 2011

[26] S. Dinha,T. Azeba,F. Fortinb,D. Mouheba,M. Debbabia, Spam campaign detection, analysis, and investigation, *Digital Investigation*, Vol. 12 supplement 1, March 2015, Pages S12–S21, March 2015

[27] C. Garrigues, N. Migas ,W. Buchanan,S. Robles,J. Borrell, Protecting mobile agents from external replay attacks, *Journal of Systems and Software*, Volume 82, Issue 2, pp. 197–206, February 2009

[28] W. Liang, W. Wang, On performance analysis of challenge/response based authentication in wireless networks, *Computer Networks* Vol. 48,no. 2, pp. 267–288, 6 June 2005

[29] D. Ruano-Ordás, J. Fdez-Glez ,F. Fdez-Riverola ,J.R. Méndez, Effective scheduling strategies for boosting performance on rule-based spam filtering frameworks, *Journal of Systems and Software*, Vol. 86, no. 12, pp. 3151–3161, December 2013.

[30] Yahoo Research, 2007, Web Spam Collections, [online], Available: http://barcelona.research.yahoo.net/webspam/datasets/, May 10, 2013.

[31] Messaging, Malware and Mobile Anti-Abuse Working Group, Report #16 – 1st Quarter 2012 through 2nd Quarter 2014,[online], Available: https://www.m3aawg.org/sites/default/files/document/M3AAWG_2012-2014Q2_Spam_Metrics_Report16.pdf, May 1, 2015

[32] G. González-Talaván, A simple, configurable SMTP anti-spam filter: Greylists, *Computers & Security*, Vol. 25, no. 3, pp. 229–236, May 2006

[33] G. Robinson, A statistical approach to the spam problem, *Linux J.*, 2003 (2003), p. 3

[34] P. hirita, J. Diederich, W. Nejdl, MailRank: using ranking for spam detection, Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM (2005), pp. 373–380
[35] G. Schryen, A formal approach towards assessing the effectiveness of anti-spam procedures, *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, 2006, HICSS'06, IEEE, , pp. 129–138, 2006
[36] L. Zhang, J. Zhu, T. Yao,An evaluation of statistical spam filtering techniques, *ACM Trans. Asian Lang. Inform. Process. (TALIP)*, 3 (2004), pp. 243–269

Authors' Profiles

Muhammad Iqbal was born in 1972 in Pakistan. He received B.Sc(Hons) and M.Sc degree in Computer Technology from Sindh University, Pakistan and MS in computer Science from SZABIST, Karachi, Pakistan. Since 2012, he is a PhD student in School of Information Sciences & Technology (SIST), Southwest Jiaotong University, Sichuan, Chengdu, PR China. His research interests are Network Security, Data Mining, Supervised Machine Learning algorithms and high speed data networks.

Malik Muneeb Abid was born in 1987 in Pakistan. He received B.Sc degree in Civil Engineering from U.E.T Taxila, Pakistan and MS degree in Transportation Engineering from NUST, Pakistan. Since 2013, he is a PhD student at School of Transportation and Logistics, Southwest Jiaotong University, Sichuan, Chengdu, PR China. His research interests are Network Robustness, Transportation network modeling and simulation, Data Mining, Supervised Machine Learning algorithms. He is member of IAROR and PEC.

Mushtaq Ahmad was born in 1986 in Pakistan. He received MS(Telecom &Networking) degree in 2013 from Bahria University Islamabad, Since 2014,he is a PhD student in School of Information Sciences & Technology (SIST), Southwest Jiaotong University, Sichuan, Chengdu, PR China. His research interests are MANETs, VANETs, Network Security and 5G WLAN Standards.

Faisal Khurshid was born in 1980 in Pakistan. He received BS(Hons) degree in Information Technology from Gomal University, Pakistan and Masters in Information Techonlgy IBMS/CS, Peshawar, Pakistan. He served as System Administartor at National University of Sciences and Tcehnlogy Islamabad from 2005 -2013. Since 2013, he is a PhD student at School of Information Science & Technology (SIST), Southwest Jiaotong University, Sichuan, Chengdu, PR China. His research interests are Network Design, Data Mining, Data Authenticity and Integrity, Supervised and Semi Supervised Machine Learning algorithms and high speed data networks.

How to cite this paper: Muhammad Iqbal, Malik Muneeb Abid, Mushtaq Ahmad, Faisal Khurshid,"Study on the Effectiveness of Spam Detection Technologies", International Journal of Information Technology and Computer Science(IJITCS), Vol.8, No.1, pp.11-21, 2016. DOI: 10.5815/ijitcs.2016.01.02