An Approach for SMS Spam Detection

Suraj J. Warade[1], Pritish A. Tijare[2], Swapnil N. Sawalkar[3]
M.E (Pursuing)[1], Associate Professor[2], Assistant Professor[3]
Computer Engineering[1], Information and Technology[2], Computer Science and Engineering[3]
Sipna C .O. E. T, Amravati, India[1, 2, 3]
surajwarade@hotmail.com[1], pritishtijare@rediffmail.com[2], swapnil.sawalkar@gmail.com[3]

Abstract- In the today's world the use of mobile increase tremendously. And hence the companies start to use of SMS for their advertisement. At the beginning the companies are send their promotional messages through SMS gateways. But due to increasing number of promotional messages the companies start the service DND, the DND service restrict only the SMS send through SMS gateways and hence the companies start to send their promotional advertisement messages through spammer's mobile phones. The approach presented in this paper detects these messages sent through spammers mobile and restrict it.

Index Terms – SMS, spam filtering, SMS spam, mobile spam.

1. INTRODUCTION

Due to daily increase in mobile users, Short message service (SMS) is very widely use text messaging service. Firstly it was designed for Global System for Mobile Communication (GSM), but now it is also available on Code Division Multiple Access (CDMA)[8]. Hence, the popularity of SMS increases day by day. The advertiser thought it is the best way to their product the one reason behind this is, first all promotional messages such as store opening announcement, shopping discounts credit card of bank etc. are sent over the email but as their spam detection facility is available all the promotional messages are going to the spam folder. And hence the companies and advertiser start sending their messages over the mobile phones as the SMS, at the beginning all messages are sent through the SMS gateways as this is easy way to broadcast message to multiple users [3]. On arrival of every message the user have to check inbox and hence because of these messages not only the mobile user is distracted but also it causes to quick fill of users inbox and user have to waste his valuable time to read and delete these unsolicited message hence, the communication service provider provides the service DND (do not disturb) which restricts the unsolicited spam messages sent through the SMS gateways [4].

When DND service restrict the messages over mobile phones. The advertiser comes with solution that send the promotional messages through the spammers mobile phone as there is not at all any restriction for messages sent through spammers mobile and it is easy and thanks to communication service provider companies unlimited SMS plans it is opportunity for spammers to use it for sending promotional unsolicited messages.

In this paper we provide the approach for detecting and restricting the spam messages sent through spammers mobile. The proposed system detect the spam messages by checking mutual relation between the sender and receiver and the content of the messages, if there is no mutual relation is found between sender and the receiver and the message contains the spam contents then the system tag the message with spam and send it to spam box and if found mutual relation and no spamming content then it directly sends to inbox of the receivers mobile.

2. LITERATURE REVIEW AND RELATED WORK

Huang Wen-Liang, Liu Yong, Zhong Zhi-Qiang, and Shen Zhong-Ming proposed a complex-network based SMS filtering algorithm which compares an SMS network with a phone call communication network. Because such comparison can provide additional features, SMS networks and obtaining well-aligned phone-calling networks that can be aligned perfectly is difficult in practice. In this paper, author presents an efficient SMS spam detection algorithm that only considers the SMS communication network. Authors first analyze characteristics of the SMS network, and then check the properties of different sets of meta-features including static feature, network features and temporal features. Authors combine these features into an SVM classification algorithm and evaluate its performance on a real SMS dataset and a video social network benchmark dataset. They also compare the SVM algorithm and KNN based algorithm to reveal the advantages of the former. Our experimental results demonstrate that SVM based on network features can get 7%-8% AUC (Area under the ROC Curve) improvement as compared to some other commonly used features [1].In [2]authors consider a local concentration based extraction approach. Two implementation strategies are designed for detecting the SMS spam as fixed length sliding window and variable length sliding window.

A novel framework for SMS spam filtering is proposed to be able to block unsolicited SMS messages by Uysal, S. Gunal, S. Ergin, E. Gunal. In the filtering framework, distinctive features representing SMS messages are identified using CHI2 and IG based features election methods. The selected features upsets with varying sizes are then fed into two different Bayesian based classification algorithms, namely the binary and probabilistic models, to classify SMS messages as either legitimate or spam. Additionally, the proposed SMS spam filtering scheme is employed to develop a real-time mobile application running on the mobile phones with Android operating system [3]. In [4] author examines the effectiveness of various content-less features that range from network and to time-oriented categories. He find that some intuitively appealing features are in fact not very effective, whereas a combination of temporal and network features can be very useful in training high performance classifiers for spammer detection.

Zi Chu, S. Gianvecchio, Haining Wang, and Sushil Jajodia focus on the classification of human, bot, and cyborg accounts on Twitter. Author first conduct a set of large-scale measurements with a collection of over 500,000 accounts. They observe the difference among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. Based on the measurement results, author proposes a classification system that includes the following four parts: an entropy-based component, a spam detection component, an account properties component, and a decision maker. It uses the combination of features extracted from an unknown user to determine the likelihood of being a cyborg, bot or human [5]. Stylistic feature that characterizes the manner in which SMS is written is introduced by authors [6]. First authors determines the style of spam messages written in Korea and they found that the most of the spam messages over Korea are sent in either English or Korean language in the same pattern and hence they provide the approach of stylistic pattern matching for detection of the spam messages. They only focus on the two languages English and Korean [6]. K. Uysal, S. Gunal, S. Ergin, and E. Sora Gunal, proposed a system for Extraction and selection on SMS spam filtering on the mobile. The author suggested the technique of data extraction of datasets like web link, alphabets, numbers, length of the message etc [7].

Artificial Immune System (IAS) of Soft Computing which Motivated by the Biological Immune System (BIS). Particularly it is based on how human immune system resist against disease and infections in the same way the mobile spam could be handled [8].Brief description about the SMS spamming methods like content matching, pattern matching and current practices for detecting spam and the data are provided in[9].

An efficient Read Aligner for next generation sequencing reads structures to detect and compare the results of web spam bot sand Viruses. This paper proposed a method of using a bio informatics pattern matching algorithm to evaluate signature-based virus/spam detection in Windows[10].

In a mobile network, viruses and malwares can cause privacy data leakage, extra charges, and remote listening. Author presented a two-layer network model for simulating and analyzing the propagation dynamics of SMS-based and BT-based viruses. This model characterizes two types of human behavior and mobile behavior, in order to observe and uncover the propagation mechanisms of mobile viruses [11].

To address the limitations of the state of research on SMS spam detection, Amir Karani and Lina Zou propose a content-based method that leverages lexical semantics. Instead of relying on individual words, proposed method uses semantic categories of words as features, which allows us to handle variations in word choices by spammers. To address the limitations of the state of research on SMS spam detection, they propose a content based method that leverages lexical semantics. Instead of relying on individual words, their proposed method uses semantic categories of words as features, which allows us to handle variations in word choices by spammers. In addition, using categories of words as features also helps to reduce the feature space, which in turn improves the efficiency of spam detection that has significant implications for SMS users. An empirical evaluation of the proposed methods has shown promising results[12].

Sr. No.	Authors	Year	Used/proposed Technique
01	Huang Wen-Liang, Liu Yong, Zhong Zhi-Qiang, and Shen Zhong-Ming	2008	Complex-network based SMS filtering algorithm
02	Y. Zhu, Y. Tan	2011	Fixed length sliding window and variable length sliding window
03	Uysal, S. Gunal, S. Ergin, E. Gunal	2012	CHI2 and IG based features election methods
04	Qian Xu, Evan Wei Xiang and Qiang Yang	2012	Non-content features (temporal and graph-topology information)
05	Zi Chu, S.	2012	Measurement and

	Gianvecchio, Haining Wang, and Sushil Jajodia,		data collection to determine user is human bot or cyborg over twitter.
06	Sarah Jane Delany, Mark Buckley, Derek Greene	2012	Study of different spam detection techniques.
07	K. Uysal, S. Gunal, S. Ergin, and E. Sora Gunal,	2012	Data extraction of datasets
08	Tarek M. Mohhamad, Ahemed M. Mahofouz	2012	Artificial Immune System (IAS)
09	D.-N.Sohn, J.-T. Lee, K.-S.Han ,and H.-C.Rim	2012	Stylistic feature
10	M. Elloumi, P. Hayati, C Iliopoulos, J.Mirza ,S. Pissis, A. Shah	2013	Bio informatics pattern matching algorithm
11	Chao Gao and Jiming Liu	2013	Two-layer network model for simulating and analyzing the propagation dynamics of SMS-based spam and viruses.
12	Amir Karani and Lina Zou	2014	Content-based method that leverages lexical semantics

Table 1: Literature Review

3. ANALYSIS OF PROBLEM

In today's world most of the people uses the mobile phone and everyone is tired of receiving, reading and deleting the spam messages. As one survey in Asia daily near about 30% of messages are spam over mobile phone in 2012[2]. Hence it is major concern about user time wastage and annoying to read and delete them.

As the mobile call and SMS log searching and analysis is most important contents to manage in soft format it needs lot storage and indexing setup. On the basis of calls and SMS log finding the direct or indirect relation between sender and receiver need lot processing which results in to non-singular record may need data analysis. Hence the proposed system is mainly designed and developed to find out the direct or indirect relation between sender and receiver to allow the textual communication. Main objective of the system is to develop SMS and call log analysis and content checking service to define sender is spam or not.

4. PROPOSED WORK AND OBJECTIVES

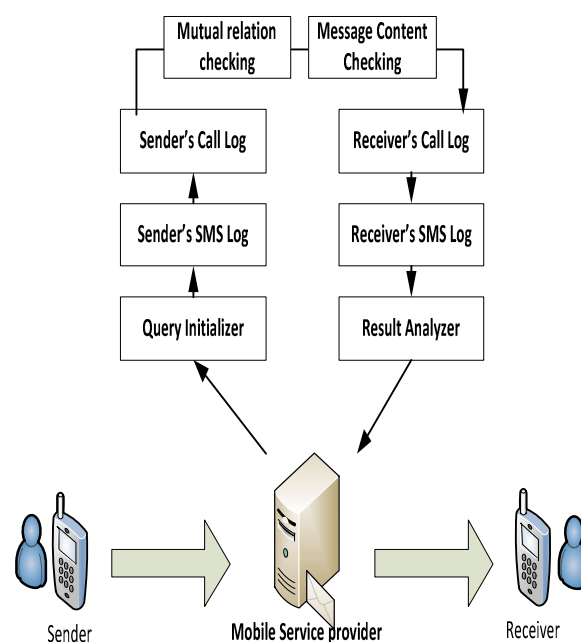Proposed system working flow is described in details in following steps,

Figure 1: System Architecture

Phase 1:
Designing and textual communication program which allows the users to register and start messaging service. It will be a prototype model of user mobile phones.

Phase 2:
Develop a server which could be a mobile service provider application server which allows the connected users to communicate. It will manage the proper transfer of messages from sender to receiver.

Phase 3:
Database designing to stores the predefined spam content & all SMS log in proper indexed format to have fast and reliable access to the records at any time of instance.

Phase 4*:*
Database analysis service which could be used to find out the direct or mutual relation between sender and receiver in order to conclude non spam entity and check for the content of message to allow the message forwarding to receiver with normal or spam tag.

In the proposed system as shown in figure above sender will first unicast, multicast a text message which will land at mobile service provider server. Once the message is received by the server then server will send the sender and the receivers address to relationships analysis module which will give the concluded result in positive or the negative format. Here the relation analysis module will look in to and previous SMS log between the sender and receiver and also look for the direct or mutual relation between sender and receiver. System will also check for the message replication or the individual message to different message and check for the content of the message. After the successful result from result analyzer system will apply and normal or spam as a tag to message and forward it to receiver or system can discard the message on the basis of configuration.

5. CONCLUSION

Mobile phone spam is a form of spam directed at the text messaging or other communications services of mobile phones. By using SMS spam detection system the system will first look up in SMS and call log data base and check a direct or the mutual relation between sender and receiver if system found no relation and if the message content are found spamming then it will treat message as a spam message and forward message with spam tag or directly reject it.

By using this system the problems occur due to spam messages like balance deduction, wastage of SMS memory etc. is get solved.

ACKNOWLEDGEMENT

REFERENCES
[1] Huang Wen-Liang, Liu Yong, Zhong Zhi-Qiang, and Shen Zhong-Ming "Complex network based SMS filtering algorithm" China Academic Journal Electronic Publishing House 13, 2008.

[2] Y. Zhu and Y. Tan, "A local-concentration-based feature extraction approach for spam filtering", IEEE Trans. on Information Forensics and Security, vol. 6, no. 2, pp. 486 – 497, 2011.

[3] Uysal, S. Gunal, S. Ergin, E. Gunal, "A Novel Framework for SMS Spam Filtering", IEEE International journal 978-1-4673-1448-0/12 2012.

[4] Qian Xu, Evan Wei Xiang and Qiang Yang "SMS Spam Detection Using Non-Content Features" IEEE Intelligent System,2012, 10.1109/MIS.2012.3

[5] Zi Chu, S. Gianvecchio, Haining Wang, and Sushil Jajodia, "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?"IEEE Transactions On Dependable And Secure Computing, Vol. 9, No. 6, November/December 2012.

[6] D.-N. Sohn, J.-T. Lee, K.-S. Han, and H.-C. Rim, "Content based mobile spam classification using stylistically motivated features," Pattern Recognition Letters, vol. 33, pp. 364–369, 2012.

[7] K. Uysal, S. Gunal, S. Ergin, and E. Sora Gunal, "Detection of sms spam messages on mobile phones," Proc. of IEEE 20th Signal Processing and Communications Applications Conference, 2012.

[8] Tarek M. Mohhamad, Ahemed M. Mahofouz "SMS Spam Filtering Technique Based on Artificial Immune System "IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012 ISSN (Online): 1694-081

[9] Sarah Jane Delany, Mark Buckley, Derek Greene "SMS Spam Filtering: Methods and Data", Expert Systems with Applications 39(10), p9899-9908, Elsevier 2012.

[10] M. Elloumi, P. Hayati, C IIiopoulos, J.Mirza ,S. Pissis, A. Shah, "Comparison for the Detection of Virus and Spam using Pattern Matching Tools", IEEE International journal ISBN:978-1-4673-5613-8 2013.

[11] Chao Gao and Jiming Liu, "Modeling and Restraining Mobile Virus Propagation", IEEE Transactions On Mobile Computing, Vol. 12, No. 3, March 2013

[12] Amir Karani and Lina Zou "Improving Static SMS Spam Detection by Using New Content-based Features", Twentieth Americas Conference on Information Systems, Savannah, 2014.