

SMS Spam Dataset Link : <http://precog.iitd.edu.in/resources.html>

Take Control of Your SMSes : Designing an Usable Spam SMS Filtering System

Kuldeep Yadav, Swetank K. Saha, Ponnurangam Kumaraguru
Indraprastha Institute of Information Technology (IIIT), Delhi
New Delhi, India
Email: kuldeep@iitd.ac.in

Rohit Kumra
Delhi Technological University (DTU)
New Delhi, India

Abstract—Short Message Service (SMS) is one of the most frequently used services in the mobile phones, next to calls. In developing countries like India, SMS is the cheapest mode of communication. The advantage of this fact is exploited by the advertising companies to reach masses. The unsolicited SMS messages (*a.k.a. spam SMS*) generates notifications, thus consuming precious user attention. To formulate spam SMS problem and understand user's needs and preceptions, we conducted an online survey with 458 participants in different cities of India. Most of the survey participants admitted that they are quite annoyed with burst of SMS spams and ineffectiveness of regulatory solutions. However, some participants reported that, they do get useful information from spam SMSes sometime (e.g. discounts at a popular food joint). In this paper, we present design and implementation of a user-centric spam SMS filtering application i.e. *SMSAssassin* that uses content based machine learning techniques with user generated features to filter unwanted SMSes and reduces the burden of notifications for a mobile user.

Keywords-SMS Spam; Spam Filtering; Machine Learning; Usable Security

I. INTRODUCTION

SMS messages are widely used by variety of services for communication such as banking updates, agricultural information, flight updates etc across the world. According to Telecom Regulatory Authority of India (TRAI), an average Indian sends approximately 29 SMSes per month.¹ Telemarketers and mobile operators have found SMS channel to be the most appropriate to reach masses in a country like India where internet connectivity is limited. In India, it is the cheapest of all current advertising solutions where one can reach to 1,00,000 people in just USD 80.² Figure 1 presents a typical promotional SMS received by a mobile user,

Most of the promotional SMSes received by the users are unwanted and considers as spam by her. In India, spam SMS flooding has become a serious problem. In 2010, total number of spam SMSes in India were estimated to be over 100 million per day³. Some of the regions in the world

¹<http://trak.in/tags/business/2009/07/07/full-report-sms-vas-usage-india/>

²<http://www.livemint.com/2010/07/27000020/Scourage-of-SMS-spam-swamps-mob.html>

³http://articles.timesofindia.indiatimes.com/2010-08-02/computing/28298744_1_mobile-users-consultation-paper-telemarketers



Figure 1: A snapshot of a received spam SMS on an Android phone

have high number of spam SMSes; for instance in 2010, about 1% of total SMSes were spam in North America whereas Asia had 30% of total SMSes as spam per day. Many countries such as US, UK, Australia have put in place strong regulatory solutions with high fine and spam reporting capabilities. India also has a regulatory solution known as National Do Not Call (NDNC) registry⁴ where one can register if she does not want to receive any mobile spam (calls / SMSes). However, NDNC is ineffective due to many reasons like lack of registrations, poor implementation due to lack of operators support and lack of spam reporting services⁵.

In terms of technological solutions, most of the work in filtering spam SMSes have been done using content of the SMSes. Similar to email spam filtering, machine learning techniques like Bayesian, SVM, n-gram have been used for classification between spam and legitimate (ham) SMS messages [5], [2], [6]. Content based filtering for SMS spam messages is challenging due to their short size, presence of large number of regional words and limited header information as compared to emails. In twitter spam detection, content information is same as SMSes but the underline social network is useful in detecting spammers as well as spam content [1].

Usually, mobile operators does not show any willingness

⁴<http://ndncregistry.gov.in/ndncregistry/index.jsp>

⁵<http://emergic.org/2010/08/09/ending-sms-spam-part-1/>

to implement a centralized spam SMS filtering solution since SMS based advertisements are part of their revenues. At the end user side, every spam SMS generates notifications and most mobile Inbox designs does not support automatic filtering or deletion of spam SMSes. Apart from SMSes, many OSNs (Facebook, Twitter) updates also got combined with messaging Inbox of the phone that generates notifications. Content on SMS channel is getting difficult to manage given the small screen of the mobile phones as well as unwanted notifications resulting in frequent disruptions for the user.

There have been some research efforts in smartly managing profile to minimize disruptions for the users but profile of a mobile is a global setting and does not provide enough flexibility to manage and customize SMS communication [3], [7], [4]. In current situation, end user is completely helpless with the available technological and regulatory solutions. Our previous work [8] have shown that decision for a being a SMS legitimate(ham) or spam is user specific. Hence, a spam SMS filtering system should be implemented at end user side to make maximum impact.

In this paper, we present design and implementation of an *user-centric* mobile based application which can filter spam SMSes or unwanted content according to user preferences. Since, content based filtering techniques have limited capabilities due to short size of SMSes, *SMSAssassin* uses content based filtering with user generated features to automatically filter spam SMSes. It uses different viewing space for different kind of SMSes to make management of SMSes easier for the user. Also, it provides an interface for customized notifications using which user can personalize to receive notifications on the reception of useful content only.

The paper is organized as follows. Section II describes application design, its different features and filtering methodology. Section III presents the description of demonstration and finally, Section IV discusses the application's capabilities, its usefulness in real-world and future work.

II. APPLICATION DESIGN

We have designed and developed an initial version of *SMSAssassin* system for filtering spam SMSes [8] which we deployed in real world for a month in 20 different participants' phone. Based on the usage patterns and user study participants feedback, we have made changes in the previous design of the application and here, we will present the new design of *SMSAssassin* application.

A. Implementation Details

We have developed *SMSAssassin* application for Android and Symbian phones both. For Symbian, mobile application code was written in Qt and Symbian C++ with around 5K lines of code and Android application was written in Java with nearly 7K lines of code. Both applications are developed as a standalone messaging Inbox application which can compose, read, delete and filter SMSes according

to user preferences. Applications have capability to read an incoming SMS and notify a mobile user based on her filtering preferences. We have developed our mobile application in such a way that it can replace the traditional Inbox of the phone with a SMS Inbox with spam filtering capability.

B. Bayesian Filtering

Bayesian filtering is an integral part of *SMSAssassin* application that is used to automatically classify SMSes with user generated features. We will discuss classification of SMSes based on user generated features in later part of this section. We used bayesian filtering in our application because it requires less computational resources and can be trained on the fly which suits for resource constrained mobile devices. Bayesian filtering is a content based supervised machine learning technique which determines whether a SMS is ham or spam based on the occurrence of keywords in the SMS. Bayesian filtering need a seed dataset containing both ham and spam SMSes to train itself.⁶ As part of training, Bayesian filtering computes count of each word's occurrence in both ham and spam SMSes (training dataset) and store them.

Whenever a new SMS comes, application extract all the words from it and compute a score with the help of training file. This score is then compared with a threshold parameter δ ⁷ to decide on a SMS is spam or ham.

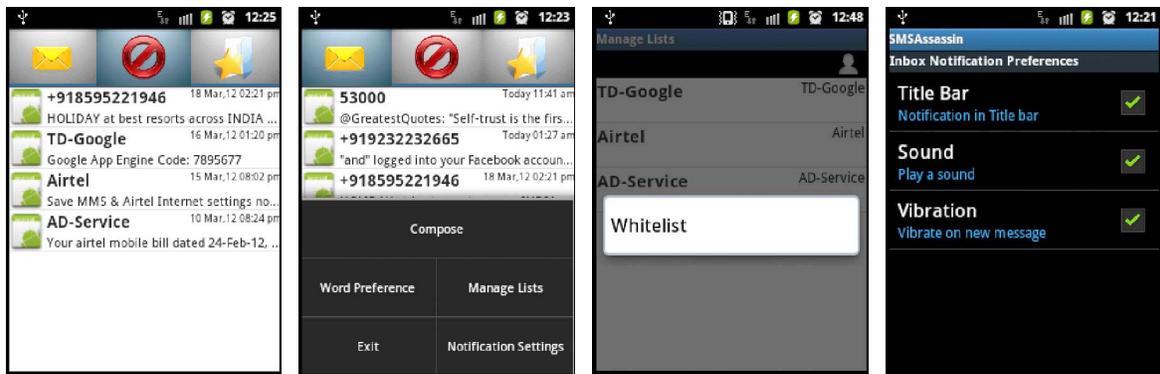
C. Features

Our mobile application works like a messaging inbox with spam filtering capability. Hereby, we will describe some of the features of the application which is not usually found in messaging inbox of the phone, some of these features also acts as classification rules and populated by users which enables personalized filtering.

- 1) **Different Tabs:** The application provides three different tabs; one each for ham (Inbox) and spam SMSes (SpamBox) respectively and third one is for user preferred SMSes as shown in Figure 2a. Whenever a new SMS comes, the application automatically decides on corresponding tab for it based on its filtering mechanism. This kind of tabular interface and aggregation of SMSes makes management and viewing of SMSes very easy for the users on small screen mobile devices. All three different tabs provide flexibility to move SMSes across different tabs to train the system in making future decisions.
- 2) **Sender Blacklisting and Sender Whitelisting :** Sender blacklisting feature is used to block a particular sender. It is a user driven feature where user blacklist

⁶We have used 1000 ham and 1000 spam SMSes for training

⁷Value of δ is determined empirically from the training dataset



(a) Main Snapshot with three different tabs for different kind of SMSes (b) Different features of the application (c) Blacklisted senders by the User (d) Notification preferences for the Inbox tab

Figure 2: Snapshots of running *SMSAssassin* mobile application in Android based Samsung Galaxy Y phone.

some of phone numbers/contacts from which she gets lot of unwanted content so that future SMSes from that sender can directly go to SpamBox. In our previous user studies [8], [9], sender blacklisting is one of highly rated feature where some of participants even blocked their contacts since they receive unwanted content from them. Figure 2c shows a snapshot of the application which shows the blacklisted senders by the user.

The application puts user's contact list into whitelist by default which can be later customized by the user. All the whitelisted sender's received SMS goes into Inbox directly. User can also whitelist a contact if it is previously blacklisted by her as shown in Figure 2c.

- 3) **Report as Ham or Spam** : If a SMS is spam and wrongly put into Inbox by the application then the user can report it as spam to tell application about wrong decision. The application gives an option to blacklist sender if a SMS is moved from Inbox to SpamBox. Similarly, if a ham SMS is put into SpamBox then user can report it as ham and subsequently move into Inbox.
- 4) **User Preferences** : In this feature, user can bind any preferred keyword (for instance, "Pizza") or phone number(s) to user preferred tab. All the SMSes which are from the user preferred sender or contains the user preferred word will come to user preferred tab (rightmost tab in Figure 2a). This feature provides an alternative viewing and storing space to the user apart from Inbox which is flooded with different kind of SMSes. In our previous user studies, most of participants demanded this kind of feature as searching of SMSes becomes difficult as number of SMSes in the messaging Inbox grows. We have seen some more

specific use cases of this feature, which is described as follows:

- a) Alice wants all the preferred promotional SMSes to keep separate from the Inbox of the phone because she uses them often to get discounts. Whenever she receives a Pizza offer or a snapdeal offer, it should directly go to user preferred tab. In the application, she would just create a word or sender filter so that the incoming SMSes comes into correct tab.
 - b) Bob has a group of friends, whenever he receives a SMS from any of them, he wants to keep it separate from the normal Inbox of the phone so that searching of a particular communication is easier and he can later refer to it quickly. In this case, he will bind phone number of all those friends to the user preferred tab of the application.
- 5) **Customized Notifications** : Notifications generated by unwanted SMS is very annoying to the users. Application provides a customized notification mechanism for each tab which can be set by user according to her own requirements, Figure 2d shows a snapshot of the application to customize notifications for all the SMSes which will come to Inbox, it has similar interface for other two tabs also. Typically, all mobile OSes provides three different types of notifications; title bar (desktop notification), ring and vibration. Our application provides interface for managing these three different kind of notification for each tab. For instance, a user may set different notification mechanisms to different tabs; she sets no notification for spam SMSes and sets high level (title bar + ring + vibration) to the a SMS coming to user preferred tab or Inbox.

- 6) **Automatic Filtering and Crowdsourcing** : If a SMS does not get classified by sender blacklisting/whitelisting or any of user generated rules then it is passed through a trained classifier based on Bayesian filtering to make a decision whether it is ham or spam. Bayesian filtering mechanism is already described in Section II-B.

Since, bayesian based filtering operates on keywords and spam SMSes keywords keeps on changing according to current trends due to different festivals etc. The application uses crowdsourcing to keep itself updated with latest trends. It logs all the spam SMSes of the mobile device, sends it to an aggregation server periodically which aggregates reported spam SMSes from all the users using the system and generates a new training files every few days which can be synced by all the mobile devices for better spam filtering.

III. DEMONSTRATION DESCRIPTION

During the demonstration, we will showcase running *SMSAssassin* application on Android phones and different features of the application described in Section II-C and their usefulness. The spam SMS filtering capability and selective notification mechanism of the application will be demonstrated by sending SMSes from other phone to the phone that is running *SMSAssassin*.

IV. DISCUSSION

Due to limited flexibility offered by current messaging inbox designs and growing unwanted content on SMS channel, end user is frequently getting disrupted. In this paper, we presented a solution to this problem by designing an application *SMSAssassin* which has potential to replace current messaging Inbox and can provide spam filtering capability thus giving control to the end user on the type of content she would like to receive. Content based filtering techniques provides limited capabilities in this problem space due to short content of SMSes and frequently used abbreviations. Our application smartly combines automatic filtering based on bayesian and user generated features such as sender blacklisting to filter unwanted SMSes and thus, saving users to frequent disruptions.

Due to lot of communication on SMS channel, the current messaging Inbox looks cluttered until user does not manually classify the SMSes and saggregate them into different categories. Also, searching of a previous communication is very difficult in an Inbox full of SMSes. *SMSAssassin* provides ways to redirect different SMS message based on their content and sender to different tabs so that they automatically gets organized and user has to do less efforts than the previous case. This kind of automatic filtering is of utmost importance since information overload on the SMS Inbox is continuously increasing due to integration of various services such as Facebook, twitter updates.

As an ongoing work, we are looking to improve our content based filtering and notification mechanism to give more flexibility to the user and minimize disruptions caused by unwanted SMSes. We will also conduct a large scale user study of proposed application to understand the application usage patterns and possible improvements in the design.

ACKNOWLEDGMENT

First author is supported by a PhD Fellowship from Microsoft Research, India. This research is partially supported by a research grant from International Development Research Centre (IDRC). Authors would like to thank Anshu Malhotra, Rushil Khurana and Dipesh Kumar Singh for their contributions in early version of *SMSAssassin*. We would also like to thank all members of PreCog and MUC research groups at IIT-Delhi for their valuable inputs and immense support.

REFERENCES

- [1] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. *CEAS 2010*.
- [2] G. V. Cormack, J. M. Gómez Hidalgo, and E. P. Sánz. Spam filtering for short messages. In *CIKM '07*, pages 313–320. ACM, 2007.
- [3] A. Dekel, D. Nacht, and S. Kirkpatrick. Minimizing mobile phone disruption via smart profile management. In *MobileHCI '09*, pages 43:1–43:5. ACM, 2009.
- [4] J. E. Fischer, C. Greenhalgh, and S. Benford. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. In *MobileHCI '11*, pages 181–190, 2011.
- [5] J. M. Gómez Hidalgo, G. C. Bringas, E. P. Sánz, and F. C. García. Content based sms spam filtering. In *DocEng '06*, pages 107–114. ACM, 2006.
- [6] D.-N. Sohn, J.-T. Lee, and H.-C. Rim. The contribution of stylistic information to content-based mobile spam filtering. In *ACL-IJCNLP Short papers*, pages 321–324.
- [7] T. Sohn, V. Setlur, K. Mori, J. J. Kaye, H. Horii, A. Battestini, R. Ballagas, C. Paretto, and M. Spasojevic. Addressing mobile information overload in the universal inbox through lenses. In *MobileHCI '10*, pages 361–364. ACM, 2010.
- [8] K. Yadav, P. Kumaraguru, A. Goyal, A. Gupta, and V. Naik. *SMSAssassin*: Crowdsourcing driven mobile-based system for SMS spam filtering. In *HotMobile 2011*, Mar. 2011.
- [9] K. Yadav, A. Malhotra, P. Kumaraguru, R. Khurana, and D. Singh. Take control over your smses: A real-world evaluation of a mobile-based spam sms filtering system. In *Manuscript submitted in APCHI'12*, 2012.