Take Control Over your SMSes: A Real-World Evaluation of a Mobile-based Spam SMS Filtering System

Kuldeep Yadav[1], Anshu Malhotra[1], Ponnurangam Kumaraguru[1], Rushil Khurana[1], and Dipesh Kr. Singh[2]

[1]Indraprastha Institute of Information Technology (IIIT), Delhi, India

[1]Tata Consultancy Services, Pune, India

ABSTRACT

In developing countries, Short Messaging Service (SMS) is one of the most widely used and cheapest modes of communication. Hence, this medium is often exploited by advertising companies to reach masses. The unsolicited (spam) SMSes consume user attention and have become a reason of annoyance for most of the mobile phone users, as not many of them use the information from these SMSes. We conducted a three phase study to understand the scale of SMS spam problem and to propose technological measures to curb it. First, we conducted a survey among 458 participants in India to understand general perceptions about spam SMSes, user preferences and requirements. Ninety seven percent of the survey participants admitted that they are quite annoyed with the burst of spam SMSes and lack of appropriate technological or regulatory solutions.

Next, we designed and implemented a mobile based application *SMSAssassin*, which can filter spam SMSes using content based machine learning techniques and user generated customized rules. And last, we conducted a user study of this application with twenty three participants who used the application in real world for a month. Results show that a mobile based solution can effectively filter spam SMSes and the application can be usable too.

Author Keywords

SMS spam, crowd sourcing, filtering, experimentation, real world

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*User-centred design*; H.1.2 Information Systems: User / Machine Systems—*Human factors*

General Terms

Design, Experimentation, Human Factors

INTRODUCTION

The number of mobile phone users in the world is over 6 billion and mobile phones are becoming ubiquitous day-by-day [5]. The unprecedented growth of mobile phones in the last two decades has opened up new ways for people to stay connected with each other. Some popular phone services include voice calls, video calls, GPRS/3G and SMSes. Globally, SMS traffic is forecasted to increase from 5 trillion in 2010 to 8.7 trillion in 2015. [1] Apart from personal communication, SMSes have become an effective communication medium to deliver banking, financial and agricultural information in developing countries like India where Internet usage is still limited. [2] Also, SMS channel has become an out of band communication channel for many other domains e.g. E-Ticketing Services. According to the Telecom Regulatory Authority of India (TRAI), an average Indian mobile phone user sends approximately 29 SMSes per month. [3]

Due to huge success and use of text messaging, SMS has now become the preferred and cheapest medium for advertisers to reach masses in developing countries like India. Using SMS, it takes only USD 80 to reach 1,00,000 people. [4] Hence, SMS channel is used by advertisers to send unsolicited (spam)[5] and unwanted content to mobile phone users. Each SMS received generates a notification in the mobile phone and most of the current phone inbox designs do not provide flexibility to delete a SMS without looking at it. Due to unnecessary notifications and content, spam SMSes have become a source of annoyance and frequent disturbance In India, during August 2010, the total number of spam SMSes were estimated to be more than 100 million per day. [6] Interestingly, the scale of spam SMSes varies across regions; for instance, in 2010 only 1% of total SMSes were spam

[1]http://www.cloudmark.com/en/spamguide/

[2]http://www.theatlantic.com/technology/archive/2011/08/why-texting-is-the-most-important-information-service-in-the-world/242951/

[3]http://trak.in/tags/business/2009/07/07/full-report-sms-vas-usage-india/

[4]http://www.livemint.com/2010/07/27000020/Scourge-of-SMS-spam-swamps-mob.html

[5]One example promotional SMS, "AVE 20-25% on Your Electricity Bills With POWER SAVER device (Govt Approved). Just Rs 999/-. Buy Today & Get a Free Holiday to Goa 4 a Couple. Call 9811641876"

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

IIIT-D Technical Report, Submitted on : 14th May 2012

in North America whereas Asia had 30% of total SMSes as spam [1]. Spam SMSes have been a problem in many countries including UK, USA, China, Vietnam, and India. Many countries including USA, UK, Australia have brought in regulatory solutions to control spam SMSes by imposing high fine and good spam reporting capabilities. India also has a regulatory solution known as National Do Not Call (NDNC) registry [7] where users can register their mobile number if they do not want to receive any mobile spam (calls / SMSes). However, NDNC service remains ineffective even after multiple attempts. [8]

The problem of desktop based spam i.e. in email [13] and social networks [4] is well studied in literature and various service providers have implemented solutions to filter it. Mobile service providers generate revenue from advertisers and thus do not take any measure to filter spam SMSes. Also, email and social media service providers have access to many other features such as subject line, long message headers, social network and past history of a sender which can be important indicators of spam.As mobiles are more handy and ubiquitous, users are expected to instantly reply incoming SMSes most of the time. Frequent spam notifications can lead to loss of attention and user may also miss out some important communication due to too many spam SMSes in the inbox. Small screen of mobile phones and single viewing interface makes content (such as SMS) management much more difficult than desktop based options. All these differences indicate that SMS spam problem is different from traditional desktop based spam and needs a new paradigm of solutions. Whether a SMS is a spam or legitimate (ham) depends on individual's choice and information needs at a particular time.

Our work [16] explores the possibility of bringing a technological intervention to fight the problem of spam SMSes keeping in mind user preferences. From the implementation point of view, there can be two types of technological solutions for filtering spam SMSes – (a) server based filtering where all the spam SMSes will be filtered by mobile operators, and the users will not receive any spam SMSes and hence will not have to pay for them (for instance, while in roaming); (b) spam SMSes will be filtered at mobile device level without giving any notifications thus saving users' precious time and attention. We believe that a filtering solution implemented at mobile device level will lead to more customization and also help address end user's privacy concerns.

Our research can be broken down into three main phases:

1. Phase 1: To understand the end user's perceptions, requirements and preferences related to spam SMSes, we conducted a survey [9] with 458 participants from India of different age-groups and cities.
2. Phase 2: We designed a system which filters spam SMSes at mobile device level. Through this system, we try to bring a fine balance between automated spam SMSes filtering and user preferences. This system is based on the work presented by Yadav et al [16].
3. Phase 3: To evaluate the system's performance in real world, we deployed this system on mobile phones of 23 users and observed their usage for a month. We analyzed mobile application usage logs and interviewed participants to gather their reactions about the application performance.

The remainder of the paper is organized as follows: In the next section, we discuss related work that helped in framing our research questions. In the User Perceptions and Requirements section, we discuss the survey that we conducted, the participants in the survey, and major findings which helped us understand spam SMS problem. In the System Description section, we present application design, implementation details, and its features. In the Application Deployment and User Study section, we discuss the user study that we conducted in real world. After that, we present the User Study results in terms of application classification accuracy and user reactions, responses and feedback. Finally in the Discussion section, we discuss the implication of our analysis on the design decisions of future spam SMS filters and Inbox designs.

RELATED WORK

We present previous work from two related directions: content based spam filtering and user studies of mobile applications.

Content Based Spam Filtering

Content based spam filtering techniques use text present in SMSes or emails to filter spam. Most prior work on spam SMSes filtering explores use of machine learning techniques to enhance classification of spam and ham (legitimate) SMSes. Most of these techniques are borrowed from email spam filtering domain where Bayesian and Support Vector Machines (SVM) based classification are the two most successful techniques [6, 9, 12, 14]. Email spam filtering systems make full use of meta data (such as header, sender, user's previous history etc.), which is unavailable in the case of SMSes, e.g. the SMS header only contains sender mobile number.

Content based filtering of spam SMSes is challenging due to their short size (only 160 characters) and presence of large number of regional words [6, 9, 16]. Techniques like Bayesian classification when used for SMSes can easily get influenced by short hand abbreviations or different variations of the same words and thus result in low classification accuracy compared to feature based techniques like SVM [1]. In contrast to purely content based approaches, Sohn et al. [14] argued that, if we consider stylistic features in SMS text then classification accuracy can be improved. Some of these stylistic features were presence of emoticons, average word length and presence of special characters.

Like SMSes, a tweet on Twitter too has limitation in terms

[7] http://ndncregistry.gov.in/ndncregistry/index.jsp

[8] http://timesofndia.indiatimes.com/business/india-business/Pesky-SMSs-down-but-not-out/articleshow/11172928.cms

[9] `http://surveymonkey.com/sms_spam_survey`

of content (140 characters). Spammers on Twitter are identified by combining content information with user attributes and underlying social network [4]. However for spam SMSes, it is hard to get user attributes and social network of the sender due to privacy restrictions and unavailability of such information.

User Studies of Mobile Applications

Most user studies focus on managing text messages or voice calls [3, 7, 15], but to the best of our knowledge, there is none on spam SMS filtering. Battestini et al. did a four month long study to understand text messaging usage in the US and found that participants engage in simultaneous conversation with as many as 9 different contacts from their mobile phone [3]. Using this, they proposed some design changes to SMS inbox like support for simultaneous conversations. Sohn et al. [15] presented a system "Lenses" which can combine messages and updates from multiple services like SMS, Facebook, Twitter and IM and extract information of interest based on sender and location. "Lenses" does not use content for filtering information due to short size of the messages. Authors in [2] distinguished SMS based communication from other desktop based communication like IM and highlighted lack of expressiveness in SMSes. They presented a system *SenseMS* which combines context information like location and emoticons to increase expressiveness of SMS communication.

Most mobile communication like calls and SMSes result in notifications and hence, users get disturbed [8, 11]. Few technologies have been built to minimize disruptions caused by mobile phones, either by smart profile management or giving user more contextual information to make an informed decision. *Telling Calls* [10] is a mobile application for managing voice calls; for instance, whenever a user receives a call, it provides contextual information like subject, location of caller, urgency, importance which can help user in taking decision whether to pick the call or not. User studies have proved that additional contextual information provided by *Telling Calls* improves the user experience and removes the necessity of other channels like SMS or IM to coordinate beforehand [10]. Dekel et al. [7] proposed a solution for smartly managing *profile* of mobile phones; they built and deployed a rule based application which can automatically change the profile based on calendar entries, location, etc. Solution proposed by Dekel et al. helps users in minimizing mobile notifications automatically during busy hours using user's calendar.

We conclude that present content based filtering approaches are less effective for spam SMSes due to short text and lack of meta data information. Also, there has been no work on minimizing disruptions caused by spam SMSes.

USERS' PERCEPTIONS AND REQUIREMENTS

There is a lack of a user-centric technological solution to spam SMS problem. [10] In this research, we try to understand users' requirements, behavior and perceptions related to spam SMSes and then explore how users respond to a personalized spam SMS filtering mobile application. Since, there is no publicly available statistics about the scale of spam SMSes and their impact on the end-users, we conducted a survey with 458 participants from different age groups and cities in India to answer the following research questions:

1. What is the current state of SMS spam problem and what is its effect on mobile phone users? Are regulatory solutions helpful in stopping spam SMSes?

2. Does the end user find any utility in such promotional SMSes (which could be spam for some)?

3. Are mobile phone users willing to use a spam SMS filtering application? If yes, what are the broad expectations from a spam SMS filtering application?

Survey

We publicized the survey through online social networks, word of mouth and distributed printed surveys to fill at public places like metro trains, shopping malls etc. Out of total 26 questions, 20 were multiple choice questions (some had a text box to provide more information), 1 open-ended subjective question and 5 questions to capture demographics. Out of total 502 survey participants who took part in the study; 458 of them completed the survey by answering all the required questions. We used only 458 completed surveys for analysis in this paper. Table 1 presents the demographics of the survey participants. We focused on getting participants across different cities / regions in India (from 60 different cities) to know the extent of SMS spam problem.

Table 1. Demographics of survey participants. (N=458).

Age	**Percentage**
15-25	59.8
26 - 35	16.8
36 - 50	16.6
> 50	6.8
Sex	
Male	63.1
Female	36.9
Occupation	
School Student	2
College Student	41.3
Government Employee	10.3
Private Sector Employee	30.8
Businessman	3.9
Non-working	7.6
Others	4.1

[10]There are some commercial spam SMSes filtering applications, for e.g. SMS Blocker (http://www.smsblocker.in/), Anti SMS for Android (https://market.android.com/details?id=org.baole.app.antismsspam&hl=en) which are based on simple rules like sender blacklisting and whitelisting to block spam SMSes. However, there is no study to measure the impact of these applications in real world. We also did not compare the filtering accuracy of our system with existing app store based applications because our solution is a combination of automatic filtering with user preferences and current solutions only uses user generated features.

Text Messaging Usage

In two separate questions, we asked participants about the number of SMSes that they send and receive per day. We asked this question because if one does not use text messaging at all, then she can stop all notifications related to messaging and avoid getting disturbed by unwanted SMSes. Table 2 presents the statistics of SMSes sent and received by respective percentage of total participants in the survey. Nearly 12% reported that they do not send any SMS but only 0.7% reported that they don't receive any SMS per day.

Some survey participants reported heavy use of text messaging. Thirty one percent of participants said that they send more than 20 SMSes per day whereas, 34% participants said they receive more than 20 SMSes per day. We found that 79% of participants out of total 142 ($458 * 31\%$) who indicated sending more than 20 SMSes per day were in the age group of 15 to 25 years and nearly 14% participants were in the age group 26 to 35 years. Also, nearly 62% participants out of 142 were students. This result supports previous research outcomes that younger people use text messaging heavily compared to older people [2, 3]. We did not see any significant difference in text messaging usage across gender.

Table 2. SMS usage among survey participants. Large proportion of participants sent or received 1 – 20 SMSes. (N = 458). Values in the table indicate the percentage of participants in each interval of number of SMSes sent and received.

Number of SMSes per day	Sent SMSes	Received SMSes
None	12.2	0.7
1-20	56.8	65.5
21-50	15.9	22.5
51-100	8.5	7.2
> 100	6.6	4.1

Mobile Spam and User Perception

In the survey, we asked the participants about the number of spam SMSes and calls they receive per day. Twenty three percent of participants reported to receive no spam calls; 65% received 1-5 spam calls; nearly 10% received 6-10 calls and rest (2.4%) reported to receive more than 10 spam calls per day. For spam SMSes, only 1.7% of the participants said that they don't receive any spam SMSes (refer Table 3). Majority of the participants (41.7% + 41.7%) received 1-15 spam SMSes per day. We found that survey participants received more spam SMSes than spam calls. Participants said both, spam calls and SMSes are equally disturbing as they both generate notifications. We did not observe any relation between number of spam SMSes received with the mobile operator. One of the participants mentioned – "Operators are selling customer leads to other business firms e.g. 'X' operator sells mobile number of their customer to Pizza Hut venture" by a working male [15-25 years].

Table 3. Average number of spam SMSes received per day. Majority of the participants (41.7% + 41.7%) received 1-15 spam SMSes per day. (N = 458).

Number of spam SMSes received per day	Percentage
None	1.7
1 - 5	41.7
6 - 15	41.7
16 - 25	9.8
> 25	5.0

Awareness and Impact of Regulatory Solutions

NDNC (National Do Not Call Registry) is a regulatory solution in India where users can register their mobile numbers to block any spam calls and spam SMSes. However, there is no publicly available information on how much impact NDNC has in stopping spam SMSes and calls. We asked survey participants whether they knew about DND (Do Not Disturb) / NDNC and whether they had registered for it. About 80% of survey participants said that they are aware of NDNC. But, surprisingly among the participants who were aware of NDNC, only about 42% said that they have registered for it, whereas 15% replied that they are not aware of registration procedure. One of the participants said – "currently, I have de-activated NDNC because job related messages [SMSes] come for interviews which are useful". Only 5 participants out of the registered 154 participants reported to receive no spam SMSes after registering in NDNC. One of the reactions we got from participants about the ineffectiveness of NDNC was – "I had activated [NDNC], but spam restarted after few spam-free weeks." In conclusion, participants felt NDNC was ineffective.

User Preferences

Survey participants were asked whether they have ever used promotional SMSes for informational or discount purposes. We gave them four options, where three of them were broad categories like food / travel discounts and the last option was that they have never used any. Seventy six percent participants said that they had never ever used promotional SMS for any useful purpose; remaining 24% participants reported to have used them for various purposes like discounts on food / travel, recharge offers, local business information and job / property / insurance related information.Thirty five percent of the users in age group 26-35 years agreed to have used promotional SMSes for the above purposes.

In a separate question we presented the participants 15 different categories of promotional SMSes which they may want to receive. Sixty four percent (294) participants selected one or more categories. The variety in the categories selected by the participants clearly indicated that different people have different information requirements and hence all promotional SMSes may not be useful for all. Also, 36% participants chose the option that they don't want to receive any promotional SMSes. This indicates the need of a personalized spam SMS filtering solution.

Table 4. Five top categories of promotional SMSes selected by male and female participants. Number with each category denotes the % of participants who selected the particular category out of total participants who selected one or more category of SMSes. N = 294, gives only for those participants, who chose to receive one or more category of promotional SMSes.

Female	Male
Food offers(e.g. pizza, coffee) - 33.1%	Food offers(e.g. pizza, coffee) - 24.9%
Entertainment (e.g. discounts on cinema) - 26.6%	Job offers - 21.1%
Job offers - 26.0%	Sports updates - 20.1%
Consumer goods (e.g. offers related to TV, AC) - 17.8%	Entertainment (e.g. discounts on cinema) - 18.7%
Education (e.g. Information on coaching) 16.6%	Banking / Insurance / Financial products /Credit cards 15.2%

User Expectations from a spam SMS Filtering Solution

We asked the survey participants about their expectations from a mobile application that can filter spam SMSes. Nearly 66% of the participants said that they will use the application if only provided for free. Whereas 14.4% of the participants said that they can pay a nominal fee for such an application; 10% of participants replied that they don't need a spam SMS filtering application; around 7% said that they have a low end and a non-programmable phone and their phone does not support mobile applications. Some reactions from participants regarding the application were – "Would love to install if it can filter Internet bulk SMSes, I've other means of getting promos but don't want at all in my phone. Basic version should be free."

To get an exact sense of requirements from the participants, we presented a set of possible features and asked to select (one or more) options which they would like to see in the mobile application. Out of 406 people who answered this question, nearly 51% answered that they would like to have automatic filtering of spam SMSes like in an email system. Forty four percent (of 406) of the participants said that they want to see separate folders for ham (Inbox) and spam (SpamBox) SMSes and nearly 43% (of 406) of the participants listed that the application should have a capability to block a sender.

Survey Takeaways

Based on survey data, we found that spam SMSes and calls are annoying to large set of people (97%) and regulatory solutions like NDNC are largely ineffective. Seventy six percent participants said that they have never used any information from spam SMSes. Even though some participants had used some promotional SMSes for getting discounts or information, all categories of spam SMSes may not be useful for all. Hence there is a need for filtering system which takes into account user preferences and information needs. Due to ineffectiveness of available solutions, mobile users showed strong willingness to use a technological solution for spam SMS filtering.

SYSTEM DESCRIPTION

In order to capture users response to a mobile based spam SMS filtering application, we designed and implemented a system called *SMSAssassin* a crowd sourcing based system for spam SMS filtering at end user level [16]. The system based on content filtering uses Bayesian algorithm and user defined preferences e.g. blacklisted and whitelisted senders to automatically determine whether an incoming SMS is a spam or a ham. This system has primarily two parts: a mobile application (running on the user's mobile device) and a server. This mobile application filters spam SMSes on the mobile phone and a server is used to crowd source spam SMSes and provide an updated training file to the mobile application for classification.

We use crowd-sourcing to learn new textual patterns which may emerge in spam SMSes during various events or festivals. For instance, we saw these SMSes during Valentines day – "Get 5% Extra discount by showing this SMS. Gift your valentine, from new Sparkles Diamond."; "Celebrate ur Valentine day with ur Family & Surprise them by Qualifying for TRIP TO THAILAND." We observed particular trends in spam SMSes for various timely events such as new year, cricket series, etc. Following subsections describe UI design and different features of the mobile application.

Application Design

Application User Interface (UI) is centered in a single screen which has tabbed interface as shown in Figure 1. We have kept application design very simple and similar to generic inbox [11] so that the user finds it intuitive to understand and easy to use. Users are able to access all features and settings of the application with a single menu based interface as shown in Figure 1. The application provides a "Hide" button so that application can run in background without disturbing usual user activities on the phone. The application provides notification to all incoming legitimate (ham) SMSes whereas there is no notification on receiving spam SMSes, thus saving users from frequent disturbances. User can check the SMSes in the SpamBox in case she wishes to.

Implementation Details

The system developed in our previous work [16] was a prototype version developed in Python for Symbian phones. For this paper, we developed SMSAssassin as a full fledged application for Symbian OS based S60 phones[12] shown in Figure 1. The mobile application code was written in Qt and Symbian C++ with around 5K lines of code. When the application starts, it imports the most recent 100 SMSes received from the phone inbox and classifies and puts them in the appropriate tabs i.e. Inbox or SpamBox using Bayesian algorithm or user defined rules. The user can now use the application like generic inbox of her phone. In future, any new incoming SMS will be classified in similar way.

[11]This term is referred for the native messaging application which comes with every phone.

[12]http://developer.qt.nokia.com/wiki/Support_for_Symbian

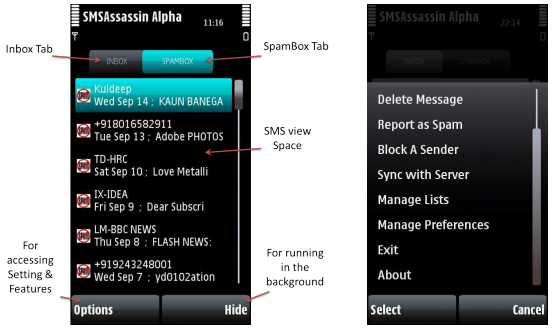

Figure 1. Snapshots of mobile application running on Nokia 5800 phones. Left: Tab view of the application with Inbox and SpamBox tabs; Right: Snapshot of the menu interface to access different features of the application.

As mentioned, the mobile application uses a combination of Bayesian algorithm and user defined preferences to distinguish between spam and ham SMSes. A user can set his preferences using different application features (explained in next subsection) e.g. sender blacklisting. This helps in customizing the application according to user requirements and helps the system in overcoming the limitations of content based techniques. Bayesian technique is used as it does not require large computational capacity from the mobile devices compared to other content based machine learning algorithms. Bayesian algorithm is trained with word frequencies which are provided to the mobile application when the application is installed. The mobile application can update these word frequencies from the server which periodically computes the latest word frequencies using spams collected through crowd sourcing.

Features

Apart from basic messaging inbox features such as reply or forward, the mobile application has following features.

1. Inbox and SpamBox: The application provides separate space for keeping ham (legitimate) and spam SMSes (see Figure 1).

2. Sender blacklisting and whitelisting: The application provides a feature by which user can block a sender by blacklisting a phone number. All future SMSes from this number will be classified as spam. Similarly, a user can whitelist mobile numbers from which she wants to receive SMSes and these will be put in the Inbox. By default, the application keeps all contacts from the phone in sender whitelist but the user has the flexibility to blacklist a contact.

3. User Word Preferences: A user can add some preferred words which will help in filtering the spam SMSes which are of interest for the user. E.g. a user may add a word preference ”pizza” to receive SMSes related to pizza in her Inbox.

4. Report as ham / Report as spam: If the application misclassifies a SMS, the user can move it to the correct tab and thus help in training the system for future decisions. Whenever user reports a misclassified SMS as ham / spam, application gives an option to whitelist / blacklist sender of that SMS respectively.

5. Sync with the server: The application logs all SMSes classified as spam by the application or reported as spam by user. By syncing with the server, users contribute spam SMSes received by her for better training of the system and downloads an updated training file of word frequencies from the server.Users have the flexibility to sync with server using her mobile Internet directly from the application or using her laptop/PC.

6. Spam deletion: The application provides an option to delete all spams at once or delete them one by one.

Whenever a SMS is received on a mobile phone, mobile application classifies and puts the SMS into SpamBox or Inbox based on the rules shown in Figure 2. We observed that usually there is a high amount of redundancy in spam SMSes received by a user. At first the application checks if the received SMS is in recently reported hams or spams. If the incoming SMS is not found in reported spam/ham SMSes, [13] the application tries to classify the SMS using user generated content. It checks if the sender of the SMS has been whitelisted or blacklisted. If the SMS is not classified according to this rule, it checks if the SMS has any of the preferred words set by the user, if so, the SMS is put in the Inbox. If the SMS is still unclassified, the incoming SMS is classified using Bayesian classification technique.

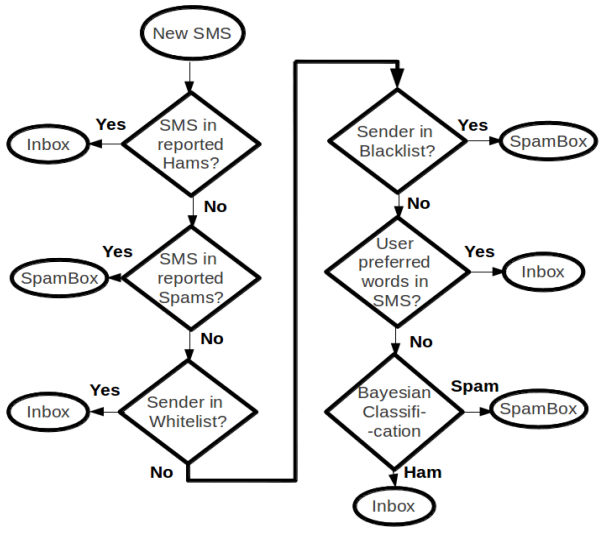

Figure 2. Flowchart of the classification rules used by the mobile application. On reception of a SMS, application decides to put the SMS in Inbox or SpamBox based on these rules.

APPLICATION DEPLOYMENT AND USER STUDY

After implementing the end-to-end system described in the last section, we deployed the system in real-world among 23

[13] We store only the last 10 reported spam and ham SMSes to minimize memory space consumption

users and monitored the application usage for four weeks. The goals of this study were: to investigate the effectiveness of the spam SMS filtering application, collect data related to typical usage patterns, observe the features most used by users, the classification rules most used by the application to make classification decisions, and get feedback and comments about the user interface and performance of the application and any further expectation from the application. All this would help uncover the design choices to be made while designing an application for mobile phones. Following research questions guided our study design :

1. How accurate is the mobile application at the end user level in terms of spam filtering? Does it vary across participants?
2. What is the effectiveness of different classification rules used by the application in real world?
3. What are the user preferences, likes, dislikes for the different features of the application and to what extent are they utilized by the users?
4. What type of user interface does a mobile user prefer for a spam SMS filtering mobile application?

Participants

Using different sources like friends, word-of-mouth, and university wide email i.e. convenience sampling, we recruited 23 participants who had a touch or a type based Nokia Symbian S60 phone e.g. Nokia C6, N5800, E50, N8. [14] We had participants from two different cities in India which helped us note the regional influence in the filtering accuracy of the mobile application. We had 18 participants from a city in Northern India and 5 participants from a city in Southern India. Participants used their own SIM and phone to use the application. To compensate participant's time and effort, we reimbursed their one month bill for using Internet on Mobile Phone. Out of the total 23 participants, 17 participants were students and 6 participants were working professionals. All student participants were in the age-group of 15-25 years and the working participants were in the age-group of 25-35 years. We had 6 female and 17 male participants.

Procedure

We met participants three times during the whole study which lasted for four weeks. During the first meeting, we deployed the application on their phones and briefed them about the application and its features. After two weeks, we met them again to see whether the mobile application was working fine. During the study if the participant faced any difficulty in using the application, they had an option to get in touch with the researchers for assistance. At the end of four weeks, we conducted a post-study qualitative interview to get feedback from the participants about the application.

Data Collection

Prior to the interview and user studies, participants were shown a printed consent form which they had to read and sign. The form stated that their application usage logs from phone would be collected, and that collected data would be anonymized and used only for the purpose of this research. Furthermore, participants were informed that they could withdraw from the study at any point of time and had the freedom to see the data being collected.

Mobile application logs: For all the users, the mobile application recorded user activity, the sender and content of spam SMSes with the timestamp. Each SMS was assigned a unique ID by the phone and using this unique ID, the mobile application logged the classification decision of the application (spam / ham) along with the rules used for each decision. In user activity, we recorded usage of all application features like report as spam/ham, word preferences, user populated sender blacklist and whitelist, movement of a sender / mobile number between blacklist and whitelist etc. The blacklisted numbers and the preferred words added were also logged. These logs were automatically uploaded on our server when the user synced with server for an updated training file.

Interviews with participants: At the end of the four weeks, we conducted qualitative interviews with all the participants. The format of interviews was semi-structured. Interview questions were primarily about the usage of different features of the application, system performance and accuracy, user interface, behavioral impact and overall satisfaction. During the interview, we encouraged participants to give suggestions and feedback to further improve the application.

USER STUDY RESULTS

Out of 23 users with whom we started the study; only 20 completed the study by using the application for 4 weeks, submitting their application logs and participating in the post study interviews. [15] In this section, we present results from analyzing the application logs and user interviews. There were a total of 5,693 ham SMSes and 925 spam SMSes classified by the application during the entire study duration. We suspect that the number of spam SMSes would have been more, but, because some people regularly delete the spam SMSes from their phone's inbox and the application is configured to import the most recent 100 SMSes only from the phone's generic inbox. Only 2 participants had less than 100 SMSes in total during the whole study. Thirteen out of total twenty participants were registered into NDNC; average number of total spam SMSes received by registered participants were 48.85 (Min = 4, Max = 114); while average number of total spam SMSes for other participants were 41.43 (Min = 11, Max = 82). It was surprising to see average number of spam SMSes for NDNC registered participants greater than non-NDNC registered participants. However, it could be a bias caused by importing of 100 SMSes from inbox or due to deletion of spam SMSes by some of the participants.

Classification Accuracy

The basic measure of performance for a spam filtering system is to see the classification accuracy in real world. We

[14] All Qt based mobile applications work on Symbian S60 or later version.

[15] One user's phone broke down during the study and two of the participants did not take part in post study interview.

logged message ID (assigned by the phone) of each SMS with the classification decision made by the application and the rule used for the decision. Though we had logged the content of spam SMSes, we did not do this for the ham SMSes in order to protect the privacy of the participants. Using the SMS ID, we kept track of the movement of wrongly classified spam and ham SMSes from the user logs and then computed the ham/spam classification accuracy for each user. For all participants, there were total 5,693 ham and 925 spam SMSes which got through our application. Out of them 178 hams were reported as spams and 91 spams were reported as hams. Using these statistics we calculated the precision, recall and overall system accuracy. Precision (for hams) was 0.968, Precision (for spams) was 0.901, Recall (for hams) was 0.983, Recall (for spams) was 0.824, and Overall System Accuracy was 0.959.

During the interview, we asked participants to give us examples of ham SMSes which got misclassified. One of the participants (No. 16) said, most of these SMSes were from *way2SMS* which is an Internet based free SMS sending service and it appends its advertisement with each SMS sent. He further clarified that, once he whitelisted *way2SMS*, everything was working fine. There may be two reasons for good ham classification accuracy; the words present in ham SMSes are typically different from spam SMSes e.g. there are more regional words and short abbreviations specific between two communicating people. Secondly, most users interact with their contacts only, so the sender whitelisting rule automatically puts the SMS into the Inbox. Some participants pointed out that typically machine generated SMSes (such as recharge notifications, banking transaction acknowledgments) were classified as spam SMSes by the application. All these SMSes had some level of text similarity with the same kind of spam SMSes which made them pass through the Bayesian classification.

Given that the interface of our application was very intuitive, participants mentioned that even moving the misclassified SMS (ham or spam) to their original location (SpamBox or Inbox) was easy. One participant said, "Selection of tab interface was good, it was very easy to check if a SMS is wrongly put into SpamBox." During the post study interview, we asked participants to rate the application accuracy on a scale of 1 to 7. All study participants provided the rating and average rating was 5.9 (var = 0.31). One male student participant commented, "the app is working good. Spam and normal messages are classified almost accurately"; another working male participant compared its accuracy with email service, "I am really impressed with the filtering of spam SMSes particularly, it was comparable to what email services provides...or may be even better."

Classification Rule Performance

The mobile application uses a set of rules for classifying ham and spam SMSes. There are five rules (Refer Figure 2): sender whitelisting, sender blacklisting, user reported ham / spam SMSes, word preferences and Bayesian filtering. We analyzed the application logs to understand the contribution of different rules in classifying the SMSes. Out of total 6,618 SMSes (combined for all the participants, 5,693 + 925), 59% of them got classified using sender whitelisting. In our classification algorithm presented in Figure 2, automatic classification technique was being applied after all other features. Due to this reason, only 24% SMSes were classified by our automatic classification technique based on Bayesian. Since participants interact mostly with people in their contacts, sender whitelisting contributed most in classification of SMSes. The application was able to classify a lot of ham SMSes with just the sender information. Various preferences set by the users had considerable impact on classification of SMSes. Among the user populated rules, sender blacklist contributed 14%, word preference list contributed 2% whereas user reported ham / spam contributed 1% in classification of SMSes. We observed that classification rule performance was different for each participant depending on her customization of various user specific features. Some participants (IDs: 3,7 and 12) had contributed least amount of user features (in terms of populating sender blacklisting/whitelisting, user preferences), however their classification accuracy was good due to automatic classification.

Table 5. Usage of different application features derived from the logs of all 20 study participants. Min and Max represent the number of times a feature was used.

User Activity	No of Participants	Average Usage Count	Min	Max	Var
Report as Ham	All	4.55	1	11	12.47
Report as Spam	All	8.9	1	30	47.57
Blacklisting Sender	All	8.90	1	29	45.67
Whitelisting Sender	19	4.47	1	8	6.92
Manage Lists	13	1.67	1	4	1.19
User Preferred Words	17	6.35	1	22	45.24
Sync with Server	15	12.13	2	44	144.98

Feature Usage by Participants

One of the main objectives of the user study was to understand the usage patterns for different features. Through the user study, we also explored the liking and disliking for different features of the study participants. Three features of the application including report as ham, spam and blacklisting sender were used by all the participants as shown in Table 5.

All participants appreciated blacklisting sender feature and termed it as one of the important feature in the application. One working male participant said, "I blacklisted the sender of spam related to share market, property as these are maximum ones [SMSes that he receives]." Across participants the average number of blacklisted senders were 8.90 (var = 45.67) which shows the high usage of this feature by the participants. Some participants blocked their contacts also; one student commented, "My contact [friend] XXXXX sends 50

SMSes a day. They went in SpamBox as I blacklisted her. I did not want to read her forwarded SMSes."

Apart from 4 participants who said that "Manage List" feature (which helped to move senders between blacklist and whitelist) is necessary in the application, all others said that they never felt a need to move a blacklisted number to whitelist and vice versa. Participants did not see any use of "Sync with Server" feature because they did not notice any difference in accuracy even after syncing with the server. However, we believe this could be due to short duration of the study and less participants, but this is an important feature to have when the spam textual patterns change over time. Three users said that they don't need word preferences feature because they don't want to receive any promotional SMSes. Seventeen participants preferred to use the feature and entered the words using which they want SMSes to be filtered. On an average, participants entered 6 words (var = 45.24) with one participant entering 22 words. Figure 3 shows the tag cloud of words entered by participants; Pizza was the most preferred category followed by Food and Movie.

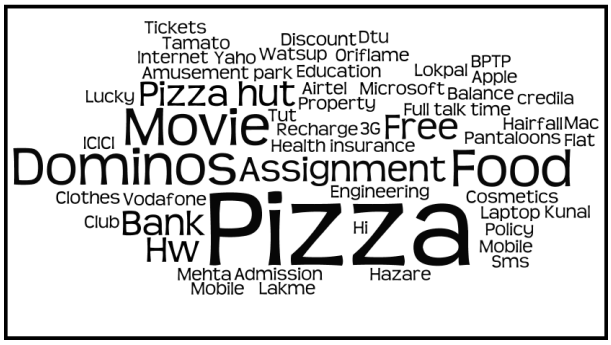

Figure 3. Tag cloud of preferred words entered by participants in the application. Pizza was the most preferred word across participants.

Likes and dislikes of Application Interface

We kept the application interface very simple. In the interview, participants gave average user interface rating as 4.82 (min = 3, max = 7, var = 1.08) on a scale of 1-7. One male student participant said, "Interface was pretty good. Light and simple. Not going for the flashy substance. Scroll was quite fast too." Some participants suggested adding some color schemes in the application to make it more attractive; "It could have been better. I don't mean that add heavy graphics and stuff, just make it a bit more playful colours."

We asked participants, would they like a tab interface or a folder interface for segregation of spam and ham SMSes. Seventeen participants preferred current tab interface of the application where one of them stated, "Tab mode is more convenient to use. I don't want to switch between folders to check if some message has been wrongly spammed [classified]." One participant mentioned "I would want to have both. Tab view to view and folder mode to segregate messages. Tab makes it easier to view, folder makes it easier to classify and store."

Overall Satisfaction

During the post study interview, we asked participants about their satisfaction in using the application. All participants provided their rating for the application on the scale of 1 to 7 and the average rating was 5.4 (min = 4, max = 7, var = 0.56). Most participants appreciated the idea of *user-centric* solution for spam SMS filtering and were satisfied after using the application. One male student participant said, "The idea of the app is really good though. Wonder why it was not there before." Another one said, "After using the application, first time I felt that I had control on SMSes which I would like to receive."

All participants agreed that the application really helped in minimizing disturbances and thus saving time and their attention. Some participants provided several other use cases where they found this application useful for their personal use. One male student commented, "I used to get really irritated especially when a string of such messages come, it annoys a lot, especially during a message conversation if you get a spam in between. But, now due to filtering of spam SMSes, I don't even bother to see them." One male participant appreciated the spam deletion feature of application, "I usually don't delete spam SMSes when they come. Due to that every week or so, I have to devote some time to delete them. Now, I can delete all spam SMSes together, it saves a lot of time for me with less disturbance during the week." Several participants agreed that due to spam SMS filtering and separate viewing space, Inbox was better organized than before.

Feedback and Suggestions

Participants suggested to add other messaging utilities like sent and draft SMSes folder and conversational view of SMSes. One male working participant suggested to add one more tab where only preferred promotional SMSes can show and another participant suggested to have a priority Inbox: "I would like some priority based Inbox to receive some important messages, only from my selected friends." Since spam calls is another medium of disturbance, majority of participants suggested to build a similar system for spam calls. Nearly all the participants agreed that application had some good features which can be useful in day to day usage and they should be part of generic inbox of the phone. Some participants complained about slow speed of the application which was primarily due to importing SMSes from generic inbox; but it can be easily resolved if spam SMS filter becomes a part of the generic inbox.

DISCUSSION

SMS is one of the most popular communication technologies among mobile phone users, but due to frequent notifications caused by unwanted SMSes, user gets disturbed. Due to frequent disruptions, mobile users tend to lose cognitive focus on the primary task [8]. The problem of spam SMSes is growing day by day and needs immediate technological intervention. We did a large scale online survey to highlight the problem of spam SMSes and end user's helplessness in stopping them.

SMSAssassin, the technological solution which we designed and developed helps filter spam SMSes at the mobile device level and minimizes notifications caused by spam SMSes. Findings from the survey pointed out that there is large diversity in people's perception on a SMS being ham or spam. *SMSAssassin* combines automatic filtering mechanism with user generated preferences and rules to filter spam SMSes. This kind of unique filtering mechanism used by *SMSAssassin* helps overcome the limitations of content based techniques and makes it easy for the user to customize the application according to its own requirements. The data collected from the semi-structured interviews and usage statistics provided insights on what application features participants would use, kind of promotional SMSes they prefer, and how the whole system performs in real world.

Our filtering system produced good classification accuracy for both ham and spam SMSes in real world deployment. Using only Bayesian filtering, Yadav et al [16] had a spam classification accuracy of 72.5% and ham classification accuracy of 97% on an India-centric dataset. Our improved average spam classification accuracy (avg = 86%) was due to user-centric features (for e.g. Report as Ham / Spam) which helped train the system according to user's need and kind of spam SMSes that the user receives. Apart from the user contacts, other user generated information (e.g. sender blacklisting) in the application helped in classifying 22% of total SMSes. It shows that users are ready to invest some time in providing training to the system, if they see a value in it. In some cases where user contributed information was limited, automatic filtering mechanism provided good filtering accuracy.

We also investigated user engagement with different features of the application from the usage logs and post study interviews. All participants liked features such as sender blacklisting / whitelisting. Interestingly, some participants blacklisted people from their contacts lists because they did not want to receive any unwanted SMSes from them. Similarly, majority of participants made full use of "word preference" feature to receive their preferred category of promotional SMSes. These findings motivate that a generic messaging inbox design should provide features for content based filtering to ensure that user is only getting content in which she is interested in.

To conclude, in this paper we found that a mobile based solution for filtering spam SMSes is effective and usable in the real world. Nearly all study participants showed their desire to use it after the study and acknowledged that application provided them control on the type of SMS content they want to receive. Users are even willing to have advanced features such as a priority Inbox and more sophisticated UI. We believe that future Inbox designs can be guided using the results from this study to help them integrate spam filtering capability as well as other features for better management of SMS communication. Recently, there has been an array of other services like IM, Social media updates integrated with the messaging inbox of the phone. Due to these services information overload on messaging inbox is increasing which needs intelligent notification management and automated content filtering at mobile device level. We believe our contributions are towards design decisions which are important while designing the mobile applications which can help user to better manage her communication with minimum disruptions.

Our study has some limitations. We selected participants using convenience sampling which is not representative of the whole population of mobile users. Even though this limits external validity of our results, we believe our conclusions are internally valid. Also, application was deployed only in two cities which was not enough to conclude regional influence in filtering accuracy of the application.

Acknowledgements

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 all the study participants for their time and feedback. We would also like to thank all members of PreCog and MUC research groups at IIIT-Delhi for their valuable inputs and immense support.

REFERENCES

1. T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami. Contributions to the study of sms spam filtering: new collection and results. In *DocEng '11*, pages 259–262.
2. A. Amin, B. Kersten, O. Kulyk, and E. Pelgrim. Sensems: a user-centred approach to enrich the messaging experience for teens by non-verbal means. In *MobileHCI '05*, New York, NY, USA, 2005. ACM.
3. A. Battestini, V. Setlur, and T. Sohn. A large scale study of text-messaging use. In *MobileHCI '10*, pages 229–238, 2010.
4. F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. *CEAS 2010*.
5. M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer. Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In *MobileHCI '11*, pages 47–56. ACM, 2011.
6. G. V. Cormack, J. M. Gómez Hidalgo, and E. P. Sánz. Spam filtering for short messages. In *CIKM '07*, pages 313–320. ACM, 2007.
7. 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.
8. 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.
9. 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.

10. S. A. Grandhi, R. Schuler, and Q. G. Jones. Telling calls: facilitating mobile phone conversation grounding and management. In *CHI '11*, pages 2153–2162, 2011.

11. J. Ho and S. S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *CHI '05*, pages 909–918. ACM.

12. M. Nuruzzaman, C. Lee, and D. Choi. Independent and personal sms spam filtering. In *Computer and Information Technology*, 2011.

13. F. Sebastiani. Machine learning in automated text categorization. *ACM Comput. Surv.*, 34:1–47, March 2002.

14. 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.

15. T. Sohn, V. Setlur, K. Mori, J. J. Kaye, H. Horii, A. Battestini, R. Ballagas, C. Paretti, and M. Spasojevic. Addressing mobile information overload in the universal inbox through lenses. In *MobileHCI '10*, pages 361–364. ACM, 2010.

16. 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.