FIMESS: Filtering Mobile External SMS Spam

Iosif Androulidakis  
MPS Jožef Stefan  
Jamova 39, SI-1000, Ljubljana  
Ljubljana, SI 1000, Slovenia  
sandro@noc.uoi.gr

Vasileios Vlachos  
Department of Computer Science and Telecommunications  
Technological Educational Institute of Larissa  
Larissa, GR 411 10, Greece  
vsvlachos@teilar.gr

Alexandros Papanikolaou  
Department of Computer Science and Telecommunications  
Technological Educational Institute of Larissa  
Larissa, GR 411 10, Greece  
alpapanik@teilar.gr

ABSTRACT

The widespread use of mobile devices has attracted the attention of cyber-criminals, who exploit their functionality for malevolent purposes. A very popular and well-known such approach is the use of unsolicited electronic messages, also known as spam. Such messages can be used by attackers in order to tempt the recipient to visit a malicious page, or to reply to a message and be charged at premium rates, or even for advertising goods and offers. Several of the mechanisms developed for fighting mobile spam have been based on the well-known and widely adopted e-mail spam techniques. Mobile spam on the other hand has specific properties, such as limited text size, particular linguistic style with specific abbreviations, also known as “the SMS language” or “textese”, etc. Our algorithm, FIMESS (Filtering Mobile External SMS Spam), performs simple, yet effective checks on the message headers so as to classify an SMS as being spam or not. In contrast to linguistic-only approaches of spam detection algorithms, FIMESS is able to utilise the important information in the SMS headers and identify SMS spam messages. Contrary to the email metadata which can easily be manipulated by the spammers, the SMS protocol provides useful information to build more efficient spam filters. The proposed scheme was tested on the Android platform and yielded encouraging results.

Categories and Subject Descriptors
C.2.0 [Computer-Communication Networks]: General—Security and protection

General Terms
Security

Keywords
Mobile spam, security, sms

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXXX-XX/XX/XX ...$15.00.

1. INTRODUCTION

Mobile phones have become a very popular means of communication. Apart from being used for telephony services, they also provide the Short Message Service (SMS) that enables subscribers to exchange short text messages among them. This service has also been exploited by businesses in an attempt to attract customers, by sending commercials or offers for discounts on certain products. When this is performed in large volumes, quite often it leads to SMS spam, a term that originated from plain text e-mails, which tried to persuade them to visit a website, in hope that they will make a purchase. In many cases, the recipient of such a message has not granted any permission to the sender for sending the message [10, 21]. The “traditional” e-mail spam problem is well known and has been attacked from many angles over the last years. Current anti-spam algorithms are utilising a wide range of techniques: human classifiers, rule-based filtering, whitelists, blacklists, collaborative spam filtering, challenge-response, greylisting, machine-learning and probabilistic classifiers, to name a few. The results of these efforts are important, as most modern filters are able to practically eliminate all e-mail spam messages with minimal False Positive and False Negative Rates (FP ≈ 0.2%, FN ≈ 2%) [12]. This paper presents FIMESS: Filtering Mobile External SMS Spam, a relatively simple, yet effective method for detecting SMS spam, upon its arrival to the mobile phone. The algorithm is particularly effective against messages originating from external mobile networks, because it analyses the technical information in the SMS headers, as well as the linguistic and the stylistic properties of the SMS. The proposed scheme has been tested on a variety of android-based smartphones with encouraging results.

The paper is organised as follows: Section 2 gives detailed facts about SMS spam, in order to set the magnitude and the seriousness of the problem, followed by examples of counter-measures that have been proposed and/or are in use to date. Section 3 describes the system’s architecture and analyses some key aspects of the proposed algorithm. Section 4 emphasises on certain key-points and includes the findings on the proposed method’s performance. Planned future work is presented in Section 5 and finally, the paper concludes in Section 6.

2. RELATED WORK

The case of SMS spam and the related countermeasures is an arms race: On the one hand, the scientific community
tries to develop techniques to detect spam in the most accurate and effective way. On the other hand, spammers do their best to avoid detection. In any case, the SMS spamming problem is quite multi-faceted, as there are several “weak links” that can be exploited from the spammer’s point of view.

A recent survey revealed that 83% of European operators do not use any sort of filtering system to protect their customers from spam, although a significant number of their customers are affected by it (around 20%). Moreover, it was also revealed that around 12% of people of the five largest European countries were receiving SMS messages from companies, without having previously granted to them any permission to do so. Furthermore, this percentage grew by 21.3% from June 2007 to June 2008 [1]. At the same time, another survey revealed that mobile operators have been struggling with growing mobile spam and malware, which has risen from approximately 2% to 20-30% of the total traffic and is expected to rise even more from year 2011 onwards [2].

The current security protocols related to SMS need to be revised, updated and offer stronger authentication, thus shielding against DoS attacks on SMS gateways. In this way, any kind of spoofing from an external party would be impossible and therefore no adversary would be able to exploit the gateway for sending spam messages in bulk. Nonetheless, solutions exist in the market for creating a secure link between the application server and the SMS gateway [23].

An SMS sent by an individual gets stored to an SMSC (Short Message Service Center) and is delivered to the recipient as soon as they are available. Hence, by using an appropriate sniffer, one can acquire the login and password information of a message sent from an SMS gateway to an SMSC. The acquired credentials can then be used for setting up a fake gateway for sending malicious or spam SMS through the SMSC. In many SMSC protocols the original sender of the message can be substituted with another one, by using a specific field. Hence, by spoofing the original sender’s number, the message will seemingly have come from a mobile phone other than the one it really has [20].

User awareness and training is –as usual– another very important factor. Educating users to guard their mobile phone numbers in the best possible way can be a very effective technique. For instance, informing them to avoid entering their number into sites offering “free” ringtones, would most probably save them from having their number being used by companies for advertising and spamming purposes.

Some carriers, such as AT&T in the US, include active users in their anti-spam measures by allowing subscribers to forward to the carrier any messages they consider them to be spam [17]. Others create alias addresses that are used instead of the phone’s real number. Messages intended for an alias address are delivered to it, otherwise they are discarded.

There are also cases where the operator offers the extreme option to completely disable all text messaging services for a given account. However, this is satisfactory for a very small percentage of their subscribers. A variant of this action is to block a subset of the text messages that satisfy certain criteria. For instance, T-Mobile, AT&T and Verizon Wireless in the US offer the option to completely block the reception of any SMS sent through the Internet.

Additional support against SMS spam includes GSMA’s spam reporting program and spam-related conventions, as as well as the Open Mobile Alliance (OMA) standards for mobile spam reporting. In the United States, recipients of SMS spam can file a complaint with the Federal Communications Commission (FCC) using Form 10881 for junk and fax telemarketing.

An on-line survey carried out in India tried to understand users’ needs and perception regarding the SMS spam problem [24]. The outcome was that users find such SMS messages annoying and that the various regulatory solutions are rather ineffective. Nevertheless, some participants seem to find some useful information in spam SMS messages, such as various kinds of discounts and offers. The authors also proposed a user-centric spam filtering application (SMSAsassin) that employs content-based machine learning techniques on the information stored on the mobile phone that are then used for identifying unwanted notifications.

Several of the schemes for countering SMS spam proposed to date use some sort of a text classification method based on machine learning. However, in order for such techniques to be effective, they first need to be trained by processing a substantial amount of text messages, before they can successfully classify messages. In turn, this brings the problem of where to store the required data. One approach is to store it on the mobile phone itself, which preserves the user’s privacy, but also introduces a computational overhead to the mobile phone. In addition, the available storage space on a mobile phone is usually less than e.g. that of a personal computer and this difference is quite significant when it comes to mid-range mobile phones, as it may have a negative impact on the system’s accuracy, due to the smaller size of the data set. Nevertheless, there are still approaches like the one proposed in [18] that feature reasonable accuracy, minimum storage consumption and acceptable processing time. Another approach is to store the data externally to the phone, process it to derive the new filter parameters and then update only the filter data on the phone.

Testing the performance of anti-SMS-spam algorithms based on machine learning requires a substantial dataset. Unfortunately, such datasets are very few and rather small in size (e.g. The DIT SMS Spam Dataset contains 1,353 unique spam SMS text messages collected from two UK public consumer complaints websites [14]). The work presented in [15] attempts to deal with the absence of large datasets and presents a large SMS spam collection, taking care to filter out duplicate entries, thus rendering it reliable to be used for evaluating and benchmarking various classifiers. At the time the article was published, the dataset contained 71,000 messages, focusing on English and Mandarin Chinese.

The authors in [13] propose a network-based online detection method for SMS spam messages. Their method employs robust text signatures and is able to identify similar messages that are sent in excess in the SMS platform. Their scheme uses counting Bloom filters to keep track of message content occurrences. Also, detection is possible in early stages and therefore the system is able to serve as an early warning system for SMS spam. Their algorithm is claimed to be quite fast and it also does not require for the messages to be stored, hence respecting users’ privacy.

The scheme proposed in [22] follows a proactive approach, where the system attempts to verify that the sender is in-

---

1 Available online at: https://esupport.fcc.gov/ccmsforms/form1088.action
ded a human and not a malicious computer program (bot). As soon as the sender attempts to send a message, he is subjected to a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) test: A picture is sent to the user in the form of an MMS (Multimedia Messaging Service), together with four possible options as to what is shown in the picture. The user must then reply and send the number of the correct answer as an SMS message. Should he fail to pass the test, the SMS will not be forwarded to its intended recipient.

The work in [16] deals with spam SMS messages that contain text in a mixture of Thai and English language words (it is worth pointing out that Thai has a completely different character set). In particular, two different filtering methods were implemented and tested. The first method uses current English message spam filtering and is then extended to support the Thai language. In the second method, the text undergoes some preprocessing that consists of normalization, word segmentation and semantic analysis/correction of the Thai words. In each method, two different decision-making algorithms were used: Support Vector Machine (SVM) and Naive Bayesian (NB). Their results revealed that the second method was almost 8.5% more accurate than the first one. An additional outcome of their experiments was that the SVM-based algorithm was approximately 2.5 times slower compared to the NB-based one, in both filtering method variants.

The work presented in [19] proposes a classification scheme that models the content and patterns of SMS syntax into an ed-weighted graph, by taking into consideration the mod-variants. Compared to the NB-based one, in both filtering method the SVM-based algorithm was approximately 2.5 times slower while now is using SMSC +3592xxxxx.

2. Usually, the first few digits of an SMSC match the respective digits of the sender’s number. This leads to the provider’s identification via the numbering plan, unless number portability is in effect. In this case, the matching digits are reduced to country prefix and possibly to mobile services prefix (e.g. +3069 for Greece, where +30 is the international prefix and 69 is the national mobile phone services prefix according to the numbering plan). As such, a message sent from an SMSC belonging to another country than the sender’s country is most probably spam, sent from some bulk SMS service situated abroad. Consequently, subscribers who reside in countries with SMSCs that routinely send spam messages can only be protected by black-listing all suspicious SMSCs.

3. The sender’s identification is checked to determine if it is a purely numerical one, or it contains other characters as well. A message with a non-numerical sender ID has a great chance of being spam. Bulk SMS providers used to allow the sender to freely choose his sender ID, greatly facilitating spoofing attacks. Since this was greatly abused, some providers enforced the mandatory inclusion of at least one nondigit character in the sender ID field. Spammers soon started using letters that resemble to numbers. For instance, the letter ‘O’ instead of the number 0 or character ‘I’ instead of number 1. Such characters combined with numerical digits in the sender’s ID are a further indication of a possible spoof attempt. Therefore, only digits and the ‘i’ sign are considered valid. Anything else will flag the message as being suspicious.

Another relevant “feature” that can be manipulated by fraudsters has to do with the way mobile phones map the numbers dialed (or the incoming numbers calling, or the ID of the sender of an SMS) to the respective entries in their catalog list. The catalog application in most cell phones matches only the 6-8 last digits of a number in order to show the respective entry name. Let’s assume that a user has an entry “Iosif” with number “+306912345678” in his phone. A spoofer sending an SMS with a sender ID of “AAA345678” will manage to get his message delivered to the victim’s phone with the display showing “Iosif”. Because of this, although the message has arrived from a completely unknown number, the display shows a contact in the user’s list, thus making a social engineering attack trivial [5].

4. The time zone of the SMSC and that of the mobile phone can also be used as possible indicators of a spam message. Our system designates a message as possibly being spam if the sender’s number is in the same country as the recipient but the time of the SMSC is more than one minute ahead. This is an indication that the message is coming from a country with a different time zone than the country of the sender’s number (a different country is what is important). Unfortunately, the
check cannot be extended to messages stamped with a time earlier in the past, since this can be appointed to delays and not necessarily a country with a different time zone.

5. Keywords blacklist. Blacklisted words such as “viagra”, “replica” etc. immediately characterise a message as being questionable.

New words can be added to the list by the user.

6. If applicable, it is checked whether the reply path field is being used and the message is classified accordingly. The specifications by the 3rd Generation Partnership Project (3GPP) allow a reply to an SMS to be sent from an SMSC other than that of the sender’s. In practice, providers do not permit this functionality, however, it could potentially be exploited to perform an attack.

7. The application checks whether HTTP links exist in the SMS and informs the user accordingly, since such a link could point to malicious content.

8. In addition, the SMS protocol ID (TP-PID) [3] is checked. There are some special-purpose SMS messages, the so-called “silent SMS” that get automatically deleted upon reception. These can be used for determining whether a mobile phone is switched on or not, without making the user aware of this action [6]. Despite the automatic deletion of the SMS, the application is able to notify the user that such an SMS was received. Furthermore, other techniques exploiting specific User Data Header (UDH) and User Data fields are also communicated to the user, since the application is checking for the presence of such information in the message [5]. Appropriate messages appear in the mobile phone’s screen, if any of the above rules is satisfied, such as the existence of invalid characters, time inaccuracies or blacklisted keywords).

4. DISCUSSION

The evaluation of the performance characteristics of every security algorithm is based on the ratio of false-positive to false-negative incidents (FP:FN ratio). Though the FP:FN ratio is widely used in epidemiological studies, is also essential for measuring the effectiveness of most security systems and applications (Intrusion Detection Systems – IDS, Intrusion Prevention Systems – IPS and AntiVirus applications). The efficacy of a security application under the FP:FN ratio can only be tested using a realistic and comprehensive dataset. Despite the fact that we evaluated all, to the best of our knowledge, available public datasets for mobile spam we didn’t find one that matches the peculiarities of our algorithm. In particular we used the following datasets:

1. The DIT SMS Spam Dataset consisting of 1,353 unique spam SMS text messages collected from two UK public consumer complaints websites [14]. This dataset is appropriate for testing the system for false negatives.

2. The NUS SMS Corpus is a well-known mobile spam data which contains perhaps the largest database of legitimate messages or hams that were voluntarily submitted for research purposes from students of the National University of Singapore [11]. The acquisition of legitimate SMS messages is a much simpler procedure, but it is necessary for determining the false positive alerts of a system. A security algorithm prone to overreaction issues is not acceptable due to the high number of the required scans on a large number of digital objects.

Our implementation, contrary to most of the related systems discussed in the Section 2, is based on the technicalities of the SMS protocol rather than the linguistic of the SMS message. As it was demonstrated in the related section 3 the algorithm is able to identify technical abnormalities and flag the suspicious messages. All the available datasets on the other hand are stripped of the required metadata and can serve only for linguistic analysis. We were thus unable to test our system in a more realistic environment with a sufficient number of SMS messages. Hence, we relied on our personal and our colleagues’ SMS messages, in order to test the effectiveness of the algorithm. None of the limited spam messages we received was able to evade detection from the spam filter nor a legitimate message was falsely identified as malicious. These preliminary tests are certainly far from reaching strong conclusions.

Although the absence of a proper dataset is the most challenging problem, other issues need to be tackled as well. A main difference between traditional e-mail spam and mobile spam is cost. Although the prices for sending SMS are low enough to allow mass distribution of spam, it is not free and therefore most spammers are more conservative in investing in similar frauds. On the contrary, e-mail spam has zero cost to the perpetrators, thus making them able to send hundreds of millions of messages. Hence, they are in a position

![SmsSD Flowchart](image-url)
Mobile spam is expected to be an important problem in the next few years, therefore appropriate solutions should be in place to minimise the volume of the unsolicited communications. The implementation of the SMS protocol itself leaves several details unresolved, which are constantly exploited by spammers to avoid detection. Despite the fact that mobile devices continuously become more powerful, they still have significant computational power restrictions. Consequently, computationally-intensive algorithms able to detect SMS spam are not suitable to be run on the mobile phone itself. Most of the current efforts to reduce the impact of mobile spam explore the linguistic characteristics of spam messages. On the other hand, FIMESS is heavily based on the technical manoeuvres that most spammers employ to deliver their messages. The proposed algorithm flags these inconsistencies in suspicious messages and, if multiple criteria are satisfied, identifies them as being spam.

### 6. CONCLUSIONS

Mobile spam is expected to be an important problem in the next few years, therefore appropriate solutions should be in place to minimise the volume of the unsolicited communications. The implementation of the SMS protocol itself leaves several details unresolved, which are constantly exploited by spammers to avoid detection. Despite the fact that mobile devices continuously become more powerful, they still have significant computational power restrictions. Consequently, computationally-intensive algorithms able to detect SMS spam are not suitable to be run on the mobile phone itself. Most of the current efforts to reduce the impact of mobile spam explore the linguistic characteristics of spam messages. On the other hand, FIMESS is heavily based on the technical manoeuvres that most spammers employ to deliver their messages. The proposed algorithm flags these inconsistencies in suspicious messages and, if multiple criteria are satisfied, identifies them as being spam.

### 7. REFERENCES

In IEEE International Conference on Communications (ICC), pages 999–1004, Ottawa, ON, Canada, 10–15 June 2012.


