Proceedings of the IASTED International Conference
Communication, Network, and Information Security
November 14-16, 2005, Phoenix, AZ, USA

LOHIT: AN ONLINE DETECTION & CONTROL SYSTEM FOR CELLULAR SMS SPAM

Siddharth Dixit Sandeep Gupta Chinya V. Ravishankar

Department of Computer Science
University of California, Riverside
Riverside 92521, USA
{sdixit, sandeep, ravi} @cs.ucr.edu

ABSTRACT

The efficient and accurate control of spams on mobile handsets is an important problem. Mobile spam incurs a cost on a per-message basis, degrades normal cellular service, and is a nuisance and breach of privacy. It is also a popular enabler of mobile fraud. In countries such as South Korea and Japan, Mobile Spamming generates almost half of the total SMS traffic.

In this paper we propose a novel spam control technique based on random projections, designed to run on SS7 links so that spams are supressed before they reach users. Our is a non Bayesian, non–keyword approach, which rate limits candidate spam messages to foil spammers.

We demonstrate using real-world spam messages that random projection is a robust, efficient and accurate method to identify SMS spam. We give a mathematical formulation of the SMS spam problem and demonstrate that it models the real world short message spam paradigm accurately. Based on this formulation, we describe a framework and algorithm to efficiently identify and control spam messages at the SMSC switch of a mobile network.

KEY WORDS
Spam, short text message (SMS), SMSC, mobile, Clustering

1. Introduction

The problem of SMS based spam messages has reached gigantic proportions and is responsible for huge financial losses for operators as well as subscribers. There are approximately 152 million mobile subscribers. In the US alone, the total number of chargeable person-to-person short messages was about 3 billion in 2004. In terms of revenue this amounts to roughly $300 million. Out of this message volume, 43% were projected to be spam amounting to losses worth 65 million for the receivers of this spam. This asserts the gravity of the cellular spam problem.

1. 1 SMS Message: An overview

SMS messages are text messages, usually 160 characters or less, sent wirelessly using the SMS (Short Messaging Service) standard. Text Messaging is convenient because it does not require computer or internet access and takes less time than making a phone call as it eliminates the protocols associated with the phone conversation. It is a convenient method for communicating quickly or when cellular network is busy or for some discrete communication. SMS has also emerged as a means to provide location sensitive information to the users.

2. SMS Spam

SMS spam is characterized by three attributes. First it is unsolicited. Second, it's generally sent in bulk. Finally, it is most often motivated by financial gain for the sender, at the cost of the receiver. For example, mobile spam often attempts to trick users into calling to premium rate numbers. The personal nature of mobile devices and availability of location and contextual information about the user has positioned SMS as an important medium for mobile advertising services. In [1], authors have predicted SMS to be the most successful mobile commerce application. Consequently, emerging phenomenon of mobile spam caused by unsolicited mobile messages or advertisements is becoming a major nuisance.

2.1 SMS vs. Email SPAM

Mobile spam problem is a much more serious problem than email spam. Mobile phones are perceived as very personal devices constantly by one's side. Also, the costs associated per SMS are significant. As opposed to email spam where the nuisance is experienced on reading it, mobile spam instantly intrudes into users' privacy by forcefully registering its arrival. People may have several email accounts, but carry only one mobile device.

SMS spam also differs from email spam in characteristic attributes. Email spam is generally identifiable by the key words used, and its structure, so that it is identifiable by

various methods. The Bayesian approach [2], derives a Bayesian confidence level for classifying an email as a spam based on the keywords it contains. The Markov Chain approach [3] considers a sequence of keywords and computes a Markovian weights for classifying an email as spam. The Support Vector Machine (SVM) method [5] is a machine learning approach which tries to classify an email as lying on one side of a plane dividing positive and negative training examples. All these techniques are based on using the keyword structure contained in spam objects. With SMS messages, however, the information available in terms of keywords and correlations amongst keyword sequences is limited by the small size of short messages. This reduces the applicability of classical email spam identification techniques such as [2], [3], [4] and [5] to our problem domain of SMS spam.

2.2 The Intuition

The attributes of SMS spam is different from that of email spam. They are short, severely limiting how variable they can be by text manipulation. Spammers often substitute some characters in short messages with similar alpha numeric or other characters (e.g. 'o' could be replaced by 0 and vice versa [Fig. 1]). However, spammers cannot create many SMS messages carrying the same content but far apart with respect to any metric that compares textual similarity, such as edit distance.

As pointed before, e-mail spam messages are not limited in such ways, since it is possible to pepper arbitrary amounts of "noise data" into email spam to create variants differing significantly by any similarity metric. Focusing primarily on this distinction we have developed a framework for spam detection of short messages.

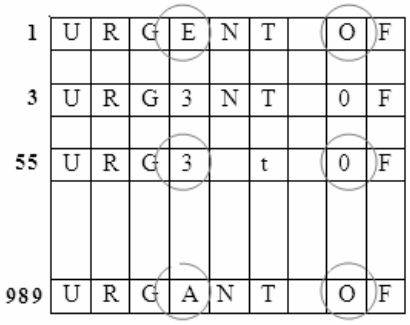

Fig 1. Cellular spam with minor modifications

2.3 Our Contribution

In our proposed method of online spam identification and filtering we have demonstrated using real work spam messages that random projection is a robust, efficient and accurate technique to identify spam short messages. We have provided a mathematical formulation of SMS spam problem and demonstrated that it quite accurately models the real world short message spam paradigm. Based on above formulation we describe a framework and algorithm to efficiently identify and control spam messages at the SMSC switch of mobile network.

3. SMS Messages: From Sender to Receiver

The point to point short message service provides a mechanism for transmitting short messages to and from mobile devices. In contrast to paging services, the service elements for SMS are designed to provide guaranteed delivery of text message to the destination.

Many cellular operators also provide the ability to send SMS messages to mobile devices via e-mail. A device's e-mail address combines its number and SMS gateway address (e.g. +1-827-234-5678@tmobilewireless.com).

Mobile-originated short messages (MO-SM) go from handset to Short Message Service Centers (SMSC) and may be directed to other mobile subscribers on mobile, fixed or electronic mail networks. Email-based SMS are routed to SMS-Gateway (SMSG) which in turn, passes them to an SMSC responsible for relaying, storing and forwarding short messages between sender and device.

The SMSC retrieves routing information by consulting the receiving device's Home Location Registry (HLR), to determine the currently serving MSC for the mobile receiving agent. It then transfers the incoming SMS to the serving MSC. The short message delivery option provides a confirmed delivery service.

SMS comprises several service elements relevant to the reception and submission of short messages

- **Validity period –** the validity period determines how long the SMSC should guarantee the storage of short message before delivery to the intended recipient

- **Priority –** Low priority messages may be delayed or starved. Our solution uses this aspect to rate control spam messages.

The architecture of ***P**ublic **L**and **M**obile **N**etwork (PLMN)* places SMSC at a vantage point through which all messages for a given PLMN, transit. This provides a unique opportunity to leverage the similarity of transitory traffic at SMSC to run spam filtering algorithm thereby stopping the cost incurred in transport of SMS over costly SS7 (signaling system no. 7) links as opposed to identifying the spam once it is delivered.

3.1 SMSC: A vantage point for SMS spam control

SMSC is a critical junction within the routing framework of short messages in the mobile network. It acts as a store and forward server and provides sufficient computing and storage capacities for spam classification and filtering. The mobile application part (MAP) [15] layer defines the operations necessary to support the SMS. MAP layer using the services of SS7 transaction capabilities part to provide transport container services for entire short message transaction. This entire transaction for a short message is visible and modifiable at the SMSC.

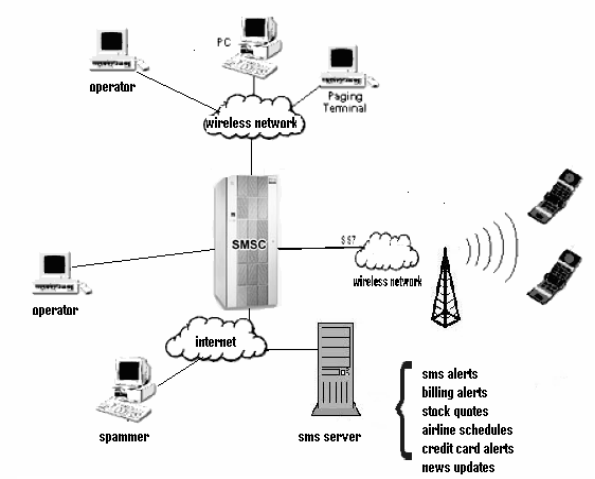

Fig. 2 SMS Network Architecture

4. Characteristics and Requirements for Spam Control

The ability to spam in an SMS area tends to be limited. SMS spammers have limited ways to change message without changing the meaning of message substantially. Also, short messages are not real-time, giving us a time window to collect and identify spam. Third, in the SMS network architecture, the SMSC is a vantage point for aggregation of all spam messages. Hence, there is no advantage to spamming from multiple sources. Next we list requirements for an SMS spam filtering algorithm.

Requirement 1*: The technique identifying spam messages should be based on the degree of similarity of short messages rather than the content (keywords) of an individual message.*

The textual and keyword based information available is insufficient to classify short messages as spam. The spam could be just 15-characters string trying to dupe a user to return a call to the spam originator, tricking him into calling a premium-rate number. Hence, filter algorithms for SMS spam must not depend on keywords or textual information but should rely upon the degree of similarity of the spam SMS being circulated in the PLMN.

Requirement 2*: The technique for spam classification should be robust to minor modifications to the subsequent short messages transiting the public land mobile network*

Spammers are constantly adapting to new techniques being developed for spam filtering. To hoodwink spam filters based on our technique spammers would aim for minor alterations which without loss of meaning of the short message. For example the spammers may aim to substitute certain characters in short messages with similar alpha numeric or other characters (e.g. 'o' could be replaced by 0 and vice versa).

Requirement 3*: The technique for spam clustering and classification should generate an extremely low rate of false positives.*

SMS messaging is used to support important services such as credit card alerts and bill payments. For this filtering scheme to be successfully incorporated for online spam filtering in cellular networks, it is essential that it should not put user at the risk of accidental deletion or delay of important SMS messages.

5. Preliminaries

We now define our setup for spam identification including notations, the premise for classifying a message as a spam, drawbacks of such an approach and naïve SMS spam identification techniques including our initial experimentation.

5.1 Setup

We treat each SMS messages as a point in *n*-dimensional Euclidean space where *n* is the maximum length of SMS messages (typically $n = 160$). Each dimension has 126 distinct values.

5.2 Notations

In our approach, Ψ_d is a high dimensional space where the original spam vectors exist, and Ψ_k is a lower dimension space. N_{spam} *and* N_{norm} are regions corresponding to spam and normal messages respectively. The SMS traffic arrival rate λ_{arr}, is sum of λ_{spam} and λ_{norm}. We use φ_r to denote clusters of radius r in Ψ. $P_s(m)$ is probability of message *m* being a member of a spam cluster and $P_{norm}(m)$,the probability of the message belonging to normal cluster. $\Gamma(t)$ is the threshold value for classifying a region as spam and $\lambda_{threshold}$ is the corresponding threshold arrival rate for spam in time interval denoted by $\Delta(t_{ival})$.

5.3 Definition of Spam

Let *r* be a constant and let S be a set of *n* SMS messages (or points) collected over time duration $\Delta(t_{ival})$. A subset S' of S is classified as spam if $|S'| > \Delta(t_{ival}) \, x \, \lambda_{threshold}$, and the corresponding points are contained within some hyper sphere of radius less than *r*. Intuitively, messages in S' form a dense cluster in an Euclidean space. For each such cluster identified by our

algorithm, we will define a confidence level, κ^+ or κ^-, of it being spam. Further, the probability of a message m being spam is $P_s(m)$.

5.4 Quantifying Spam

The confidence level for categorizing message(s) as spam is derived from the spam identification run "at hand" and previous run. In this section we define a quantitative definition of confidence level for classifying a cluster and an individual message as spam. We define $P_s(m)$ as the probability of message being spam and $P_{norm}(m)$ the probability of message being a normal message. $P_s(m)$ is $\propto 1/\mathbb{R}_m$ where $\mathbb{R}_m$ is distance of the point representing short message m from the centroid of the cluster. Also $P_s(m) \propto \kappa$ where κ is the confidence assigned to a region to be spam based on the following case:-

- The region is assigned positive confidence level κ^+ if it overlaps the spam region(s) N_{spam} identified in the previous runs.
- The region is assigned a negative confidence level κ^- if the region under consideration overlaps the misclassified spam (normal message clusters) region N_{norm} identified in space Ψ_d during previous runs. Such regions are would be observed in case of bulk but legitimate text messaging.

$$P_s(m) \propto \begin{cases} \kappa / \mathbb{R}_m \text{ if } |S'| \geq \Delta(t_{ival}) x\ \lambda_{threshold} \\ 0 \text{ otherwise} \end{cases}$$

Thus with probability $P_s(m)$ we could declare a message m to be a spam based on the fact that (a) it belongs to the cluster of sufficiently high spam message density and (b) this region was previously declared as a spam region with confidence κ. This definition lends itself naturally to our assertion that clusters of high message density are most likely candidates of short message spam.

5.4 Drawbacks

The main drawback of this definition is that it does not automatically differentiate the clusters of spam and legitimate messages. To overcome this problem we assign a confidence κ^- to the clusters above threshold which have been found to lie in regions corresponding to legitimate messages clusters identified during previous runs or learning period.

6. LOHIT

Based on the characteristic requirements and definition of SMS spam, we describe, LOHIT, a framework for cellular spam detection and filtering.

6.1 Architecture for LOHIT

Our system identifies spam at its arrival at SMSC. The signalling links of a mobile network carry short text messages to SMSC which as detailed, acts as a store and forward switch. We define the architecture of Lohit with respect to spam identification process at SMSC as follows

- the SMSC acts as a store and forward switch, this provides a time window $\Delta(t_{ival})$ to run a spam identification and control algorithm.
- Upon arrival of short text messages for a given time window, messages are converted into vectors, each representing a point in n dimensions.
- Subsequently message vectors are processed by random projection algorithm as defined in section 6.4 to reduce the dimensionality of the message vectors.
- K-means clustering algorithm is then applied to cluster the dimensionality reduced short message vectors.
- The resulting clusters are classified as spam or non spam based on criteria in section 5.4.
- Once the subset S' of messages is identified as spam, then the system aims at penalizing the spam candidates. This is achieved by either nullifying the validity period of the spam short messages or by downgrading the priority of spam messages as illustrated in section 3.
- Once the processing for a given time interval $\Delta(t_{ival})$, is completed, the system retains user tunable fraction $f_{\%}$ of spam messages vectors from the previous run for reinforced classification of similar spam in subsequent runs. A higher priority is assigned to spam messages falling in previous regions N_{spam}^{prev} of prior runs.

6.2 Naïve SMS Spam Identification

The simplest approach to SMS spam identification would be by application of naïve Bayesian identifier. Naïve Bayesian approach is applicable in case of email spam because of the very nature of email that it contains a large volume of keyword based information. This information could be correlated to known datasets of spam and a

successful and near accurate classification could be carried out. In case of SMS spam the dearth of keyword based information makes it impossible to achieve a high accuracy using such conventional techniques. Thus we resort to clustering based approach. The conventional Clustering based techniques such as SVD [13] are good techniques to cluster the short message dataset but are resource intensive techniques, unsuitable for online spam clustering and detection. To validate our hypothesis we have used SVD in initial phase to cluster the spam and project it onto 2D & 3D subspace. The results of the experimentation are shown in Fig 3(a) and Fig 3(b). In next section LOHIT we describe a framework for spam identification which harnesses the capabilities of random projection to deliver a computationally feasible solution for online operation.

6.3 Identifying spam: A random projection approach

Its been a well established fact ([6], [8] and [9]) that if dataset is clustered in high dimension then with high probability it is clustered along any random subspace. Additionally it is significantly less expensive to compute than other methods of dimensionality reduction such as PCA [14]. Later, through experiments show that our approach meets all the requirements for spam filtering mentioned previously.

Therefore we propose to cluster the points gathered in one round at SMSC by first projecting the points onto randomly selected low dimensional space. Once we have our data set in a low dimensionality space and suitable clustering techniques could be applied to classify the region as spam or normal. Next we outline our procedure in greater detail.

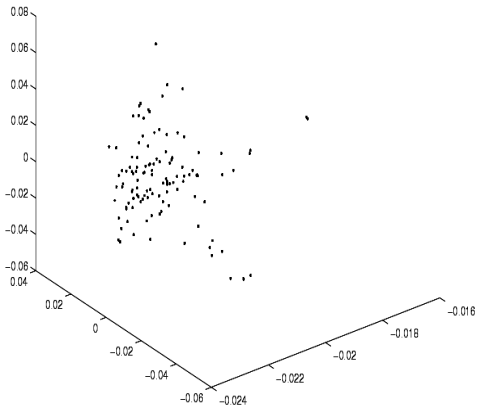

Fig 3(a) spam messages in 3D subspace

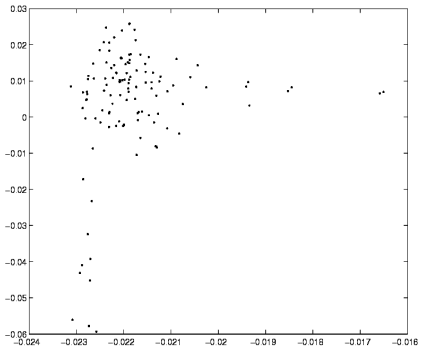

Fig 3(b) spam messages in 2D subspace

6.4 Procedure for Projection onto Random Subspace

In the next few paragraphs we describe our approach consisting of dimensionality reduction (Random Projection), Ortho-normalization (Gram Schmidt Process) and Clustering (K-Means).

6.4.1 Random Projection

In this method the original d-dimensional SMS data is projected to a k-dimensional (k << d) subspaces using a random k X d matrix M_{dxN}^{sms} whose rows represent short text messages, as a set of N d-dimensional vectors .

$$X_{kxN}^{RP} = R_{kxd} M_{dxN}^{sms}$$

The resultant matrix X_{kxN}^{RP} is a random projection of short text messages on a lower dimensional (k) subspace.

Next we find ortho-normalized set of vectors that span the projected low-dimensional space by using Gram-Schmidt process [10]. The ortho-normalization is carried to make the distance computation among vectors efficient. It also allows us to validate that the dimensionality of the projected subspace is close to d.

6.4.3 K-Means clustering

K-means clustering (McQueen, 1967) defined in [12] is a method commonly used to automatically partition a data set into k groups. Ei will represent message vectors after RP and Cj is a centroid of cluster j. It proceeds by selecting k initial cluster centers and then iteratively refining them as follows:

1. Each instance Ei is assigned to its closest cluster center.

2. Each cluster center Cj is updated to be the mean of its constituent instances.

The algorithm converges when there is no further change in assignment of instances to clusters.

In our work we use the k means clustering algorithm to run it on dataset comprising of short message vectors which have already been normalized by applying dimensionality reduction using RP [6] and GS [10] Algorithms.

The clusters are then marked as spam or normal based on the our definition of spam.

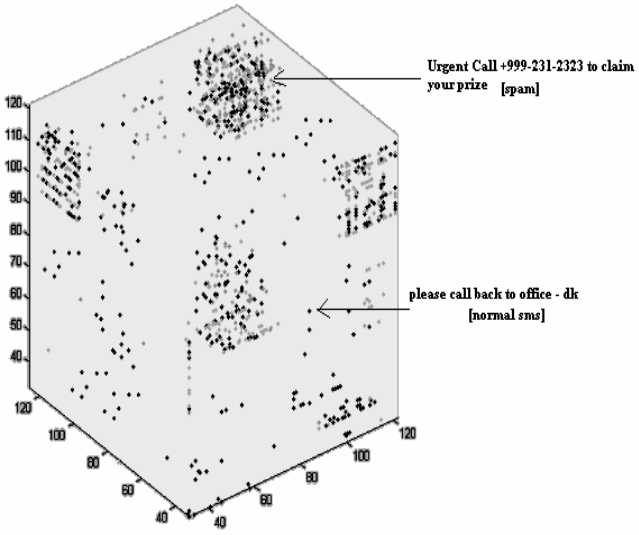

Fig 4. Short message clusters with spam and normal messages

6.5 Identifying dense spatial regions in the Native space

We cluster points in the low dimensional projected space. Cluster thus formed may or may not be the clusters in high dimensional space. Although following is true: If there is cluster in high dimensional space then it also visible as cluster in any low dimensional subspace.
Therefore to guard against false hits of spam messages we propose to run t instances of this algorithm. We say that a region is dense if its projection is identified as dense in all the instances. Based on the accuracy we observed in our experiments for spam identification by our random projection algorithm we conjecture that a 2-3 instances of the algorithm will be sufficient**.**

6.5 Demonstration over real world dataset

In our experimentation we applied the above mentioned technique to dataset consisting of real world short messages with spam short messages obtained from [7]. The dataset consisting of approximately 10,000 spam as well as legitimate short messages.

In figure 4 we could clearly see the clusters of spam messages shown as high density regions and legitimate messages spread evenly elsewhere. The clusters represent messages which are similar to each other and are at close Euclidean distances.

6.6 Rate of Arrival of Spam Vs Non Spam

We consider SMSC as our observation point which is also a transfer point for all users using short message service of the network. The link carrying messages to the SMSC and leaving the SMSC is shared by all users The users including spammer send short messages over cellular links which enter SMSC and from there messages are transmitted to the destination clients. [11] states that SMS message generation follows Poisson's distribution.

Incorporating the fact that currently a high number of short messages traversing the cellular network are spam , it's reasonable to assume one fourth of all messages to be spam . If λ_{arr} is the arrival rate and λ_{spam} & λ_{norm} are arrival rates for spam and legitimate email then in our model we assume $\lambda_{norm} \cong 3 * \lambda_{spam}$. The messages for a given interval $\Delta(t_{ival})$ are assumed to be interleaved randomly. The main goal of this model is to capture real world short message communication pattern into our experimental setup.

7. Related work

SMS spam is a relatively new but exponentially growing problem. The gravity of the problem could be gauged from the fact that mobile spam is growing at a rate of 43% in US. Still not much work has been done in field of preventing mobile spam. To the best of our knowledge this is first of its kind of work in this area. A large volume of work has been done in area of Email spam filtering. Various spam filtering techniques have been proposed for email spam. These include Naïve Bayesian Classification [2], Memory based approaches [3], Markov chains [4] and support vector machines (SVMs) [5]. Bayesian Email spam classification mechanisms claim a very high level of accuracy for email spam but are not well suited for SMS spam which is characterized by the attributes of short message length (fewer keywords) and continuously changing SMS spam messages from time to time. The most notable observation about SMS spam messages being circulated across different providers is that for a given period of time they fall under same category. Our technique is a novel approach to spam filtering of short text messages which makes use of similarity of short text messages transiting a PLMN during a time frame , to cluster them into groups of spam and seeks to rate limit the spam messages to thwart the spammers. In next two paragraphs we are illustrating some techniques used in different fields for clustering of short text strings.

8. Conclusion and Future Work

In this paper we have proposed a new algorithm for spam detection of cellular short text messages based on structural similarity of the SMS messages transiting the cellular network. The idea is that the structural similarity of transitory mobile short messages could be harnessed to

cluster the short messages and larger clusters are indicative of bulk messaging practiced by spammers. The major drawback of our algorithm is that it could cluster short messages based on their similarity in higher dimensionality space, but has no way of determining by itself if such vector clusters correspond to spam or legitimate SMS messages which are bulk send. Because of this feature this algorithm should be complemented by a learning phase where the system is trained to differentiate between spam and legitimate emails clusters. One such approach is to mark regions corresponding to spam and also to retain most frequent spam messages for subsequent runs. The regions with high density clusters, which contain known spam messages from previous runs, would be assigned higher confidence thus classifying them as spam. Given the high dimensionality short messages we have successfully demonstrated that dimensionality reduction techniques could be utilized successfully for implementing a real time software for online detection of cellular short message spam. We have tested our results on real world spam data collected from [7] and demonstrated that our technique is a viable solution for online spam filtering. It has been able to produce results of the level of the classical spam classifiers for email spam. In the light of the fact that classical methods could not be applied to cellular spam problem, our work, to the best of our knowledge is first such contribution for addressing the problem of cellular spam.

There are several improvements and developments which we have not explored but could lead to promising developments. The notable amongst them is using classical spam classifiers such as Bayesian to be applied off line on cellular spam data classified by our method to achieve a finer level of granularity of spam identification. We observe that the similarity of the short messages is a very promising attribute and gives us a basis for clustering of cellular spam. Basis of our approach lends itself naturally to the nature of problem under consideration. Finally we look this work as the harbinger in the field of cellular spam detection and anticipate it to evolve in conjunction of other techniques for collaborative spam detection of mobile spam.

References

[1] Giovanni Camponovo, Davide Cerutti , The spam issue in mobile business a comparative regulatory overview , Proceedings of the Third International Conference on Mobile Business, M-Business 2004

[2] Androutsopoulos ,. An Evaluation of Naive Bayesian Anti-Spam Filtering. Proceedings of the workshop on Machine Learning in the New Information Age, Barcelona, Spain, pp. 9-17, 2000

[3] G. Sakkis, I. Androutsopoulos, G. Paliouras, V. Karkaletsis, C.D. Spyropoulos and P. Stamatopoulos, A Memory-Based Approach to Anti-Spam Filtering for Mailing Lists ,. Information Retrieval, v. 6, n. 1, pp. 49-73, 2003

[4] Yerazunis, B.: The Spam Filtering Plateau at 99.9% Accuracy and How to Get Past It. In: MIT Spam Conference. (2004)

[5] Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In N´edellec, C., Rouveirol, C., eds.: Proceedings of ECML-98, 10th European Conference on Machine Learning, Chemnitz, DE, Springer Verlag, Heidelberg, DE (1998) 137–142

[6] E. Bingham and H. Mannila. Random projection in dimensionality reduction: applications to image and text data. In SIGKDD'01, 2001.

[7] Grumbletext.co.uk: Spam Short Messages Archive http://www.grumbletext.co.uk/vt.php?t=333&subj=complaints++SMS+Spam+%28General%29+complaint

[8] R. Weber, H.-J. Schek, and S. Blott. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In VLDB'98, 1998.

[9] Fern, X. Z., & Brodley, C. E. (2003). Random projection for high dimensional data clustering: A cluster ensemble approach. ICML.

[10] W. Hoffman. Iterative algorithms for Gram-Schmidt orthogonalization. Computing, 41:335{348, 1989.

[11] Zohar Nahor, an efficient short message transmission in cellular networks, Infocomm 2004

[12] Jain, A. K. , Dubes , R. C. 1988. Algorithms for Clustering Data. Prentice-Hall advanced reference series. Prentice-Hall, Inc., Upper Saddle River, NJ.

[13] V.C. Klema and A.J. Laub, "The Singular Value Decomposition: Its Computation and Some Applications," in *IEEE Transactions on Automatic Control*, 25(2), 164-176, 1990

[14] S. Roweis. EM algorithms for PCA and SPCA. In Neural Information Processing Systems 10, pages 626–632, 1997

[15] G. Peersman, S. Cvetkovic, P. Griffiths and H. Spear. The Global System for Mobile Communications Short Message Service. IEEE Personal Communications, June 2000.