Sindhura Parvathaneni / International Journal of Engineering Research and Applications (IJERA)
ISSN: 2248-9622 www.ijera.com
Vol. 1, Issue 3, pp.427-431

Collaborative Spam Filtering

Sindhura Parvathaneni
Mtech (SE)
Godavari Institute of Engineering and Technology
Rajahmundry

ABSTRACT

In this paper we fully deal about the concept of enormous spam e-mails directed at large numbers of recipients, designing effective collaborative anti-spam systems raises several important research challenges. Since e-mails may contain confidential information, any collaborative anti-spam approach has to guarantee strong privacy protection to the participating entities. Second, the continuously evolving nature of spam demands the collaborative techniques to be resilient to various kinds of camouflage attacks. Third, the collaboration has to be lightweight, efficient, and scalable. Toward addressing these challenges, to achieve all the above statements we create a frame work called as privacy-aware framework for collaborative spam filtering) through which we control the spam attacks. In designing the ALPACAS framework, we make two unique contributions. The first is a feature-preserving message transformation technique that is highly resilient against the latest kinds of spam attacks. The second is a privacy-preserving protocol that provides enhanced privacy guarantees to the participating entities.

Keywords: **collaboration, spam, framework, privacy, filtering**

I. INTRODUCTION

STATISTICAL filtering (especially Bayesian filtering) has long been a popular anti-spam approach, but spam continues to be a serious problem to the Internet society. Recent spam attacks expose strong challenges to the statistical filters, which highlights the need for a new anti-spam approach. The economics of spam dictates that the spammer has to target several recipients with identical or similar e-mail messages. This makes collaborative spam filtering a natural defense paradigm, wherein a set of e-mail clients share their knowledge about recently received spam e-mails, providing a highly effective defense against a substantial fraction of spam attacks. Also, knowledge sharing can significantly alleviate the burdens of frequent training stand-alone spam filters.

However, any large-scale collaborative anti-spam approach is faced with a fundamental and important challenge, namely ensuring the privacy of the e-mails among untrusted e-mail entities. Different from the e-mail service providers such as Gmail or Yahoo mail, which utilizes spam orham (non-spam) classifications from all its users to classify new messages, privacy is a major concern for cross-enterprise collaboration, especially in a large scale. The idea of collaboration implies that the participating users and e-mail servers have to share and exchange information about the e-mails (including the classification result). However, e-mails are generally considered as private communication between the senders and the recipients, and they often contain personal and confidential information. Therefore, users and organizations are not comfortable sharing information about their e-mails until and unless they are assured that no one else (human or\ machine) would become aware of the actual contents of their e-mails. This genuine concern for privacy has deterred users and organizations from participating in any large-scale collaborative spam filtering effort.

To protect e-mail privacy, digest approach has been proposed in the collaborative anti-spam systems to both provide encryption for the e-mail messages and obtain useful information (fingerprint) from spam e-mail. Ideally, the digest calculation has to be a one-way function such that it should be computationally hard to generate the corresponding e-mail message. It should embody the textual features of the e-mail message such that if two e-mails have similar syntactic structure, then their fingerprints should also be similar. A few distributed spam identification schemes, such as Distributed Checksum Clearinghouse (DCC) and Vipul's Razor have different ways to generate fingerprints. However, these systems are not sufficient to handle two security threats:

1) Privacy breach as discussed in detail in Section 2 and
2) Camouflage attacks, such as character replacement and good word appendant, make it hard to generate the same e-mail fingerprints for highly similar spam e-mails.

To simultaneously achieve the conflicting goals of ensuring the privacy of the participating entities and effectively and resiliently harnessing the power of collaboration for countering spam, we design a particular

framework and name it "A Large-scale Privacy-Aware Collaborative Anti-spam System" (ALPACAS).

In designing the ALPACAS framework, this paper makes two unique contributions:

1) We present a resilient fingerprint generation technique called feature-preserving transformation
that effectively captures the similarity information of the e-mails into their respective encodings, so that it is possible to perform fast and accurate similarity comparisons without the actual contents of the e-mails. Further, this technique also ensures that it is computationally infeasible to reverse-engineer the contents of an e-mail from its encoding.

2) For further enforcing the privacy protection, a privacy preserving protocol is designed to control the amount of information to be shared among the collaborating entities and the manner in which the sharing is done. We evaluate the proposed mechanisms through series of experiments on a real e-mail corpus. The results demonstrate that the ALPACAS framework has a comparable overall filtering accuracy to the traditional stand-alone statistical filters. Furthermore, ALPACAS resists various kinds of spam attacks effectively. For good word attack, ALPACAS has 10 times better false negative rates than both DCC and Bog filter, a well-known Bayesian-based spam filter. For character replacement attack, ALPACAS shows a 30 times better false negative rate than DCC and 9 times better false negative rate than Bog filter. ALPACAS also provide strong privacy protection. The probability of a ham message to be guessed correctly by a remote collaborating peer is well controlled below 0.001.

A. Limitations of Statistical Filtering Techniques:

Statistical filtering is currently the predominant anti-spam approach. The central idea of all statistical filters is to assign each word (more generally token) with a spam likelihood value and a ham likelihood value and classify e-mails based on the likelihood values of the words appearing in them. Naive Bayesian classifier, which is a popular machine earning-based statistical filter, generates the spam and ham likelihood values of the tokens based on the statistics of their appearances in a set of training data. For each newly arriving message, this technique calculates a score based on the spam and ham likelihood values of its tokens, which is
then used for classifying the message. With significant amount of research efforts devoted to improving its accuracy, statistical filters have been reasonably successful in filtering traditional types of spam messages when they are trained with sufficient data.

However, these stand-alone statistical filters suffer from two major limitations. First, statistical filters are highly vulnerable to a class of attacks that are intended to confuse them by appending ham-like material or reducing the spam words in the e-mails. For example, in the good word attack, the spammer appends large numbers of good words (those that appear mostly in ham messages) to the end of spam e-mails, thereby misleading the statistical filters to classify them as ham. Similarly, Picospams are extremely small e-mail messages, and they hardly contain any word that can be used by statistical filters for classification. Our experiments show that the effectiveness of the Bayesian filter can deteriorate by a staggering 55 percent, when only 20 percent good words are appended to the e-mail. Second, most statistical filters suffer from limited training set. Since the training sets are the basis upon which the spam and ham likelihood values are computed, the statistical filters are very sensitive to the accuracy and completeness of the training sets. While reasonable spam data sets are publicly available, a privacy concern has deterred users and organizations from participating in any ham archiving efforts. Thus, the sizes of publicly available ham data sets are small fractions of their spam counterparts. Moreover, in order to deal with constantly evolving spam mechanisms, statistical techniques need continuous streams of ham and spam training sets, which are generally not available. These factors have adversely affected the classification accuracies of statistical filters.

B. State of the Art in Collaborative Anti-Spam Systems:

Prior efforts on coordinated real-time spam blocking include DCC, Vipul's Razor , SpamNet, P2P spam filtering, and SpamWatch .We discuss
the drawbacks of the existing collaborative anti-spam schemes using DCC as a representative example.

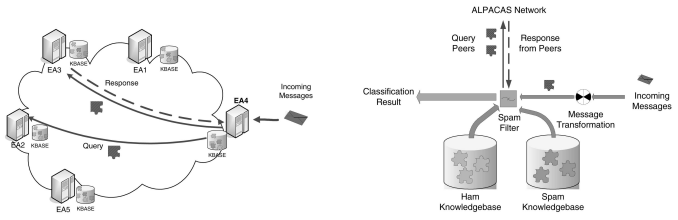
The DCC system attempts to address the privacy issue by using hash functions. Here, the participating servers do not share the actual e-mails they have received and classified. Rather they share the e-mails' digests, which are computed through hashing functions such as MD5 over the e-mail body. When an e-mail arrives at a mail server, it queries the DCC system with the message digest. The DCC system replies back with the recent statistics about the digest (such as the number of instances of this digest being reported as spam). DCC suffers from two major drawbacks: First, since hashing schemes like MD5 generate completely different hash values even if the message is altered by a single byte, the DCC scheme is successful only if exactly the same e-mail is received at multiple collaborative servers.DCC develops fuzzy checksums to improve the robustness by selecting parts of the messages based on a predefined dictionary. However, spammers can get around this
technique by attaching a few different words to each e-mail. Second, the DCC scheme does not completely address the privacy issue. A closer examination reveals that the confidentiality of the e-mails can be compromised during the collaboration process of DCC. Thus, it violates the privacy requirement from the e-mail sender for maintaining
the confidentiality of the recipients when he wants to deliver e-mails to multiple recipients by using "Bcc:." In particular, one DCC server can possibly infer who else receives the same e-mail by comparing the querying fuzzy checksum. Assuming

DCC uses perfect hash function, consider the scenario wherein an e-mail server EAi received a ham e-mail Ma. Suppose another e-mail server, say EAj, receives an identical e-mail later and sends its fuzzy checksum to EAi. Since EAi had seen this e-mail before, it immediately discovers that EAj has also received the same e-mail Ma. We refer to this type of privacy compromise as inference-based privacy breaches.

These two drawbacks, namely vulnerability toward camouflage attacks and potential risk of privacy breaches, highlight the need for better collaborative mechanisms that are not only resilient toward minor differences among messages but are also robust against inference-based privacy compromises. Anti-spam collaboration has also been proposed in the form of spam detection using e-mail social network. These approaches are orthogonal to the work presented in this paper and can be used to further enhance the effectiveness of our system.

C. Privacy-Aware Data Management:

Recently, there has been considerable research on privacy and trust issues in data management Data perturbation and data anonymization,are the two basic approaches for ensuring privacy of relational data. Researchers have also proposed various privacy-aware schemes for sharing information among independent databases. Further, the problems of privacy-preserving query computation and data mining have also received considerable research attention. However, most of these schemes cannot be used for collaborative spam filtering application as the underlying
data is essentially textual in nature.



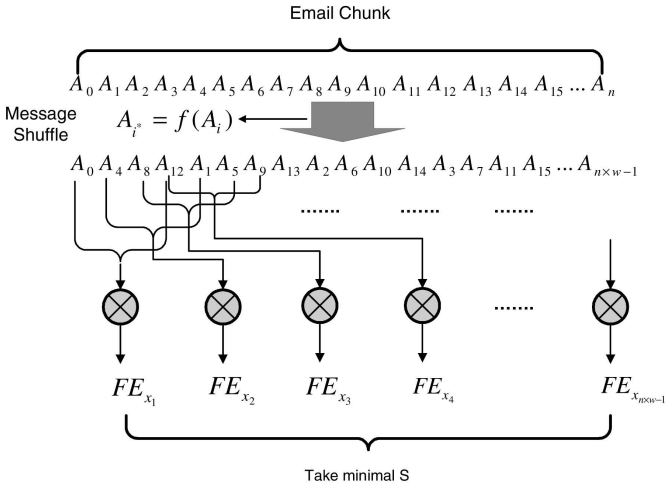
II.THE ALPACAS ANTI-SPAM FRAMEWORK:

We present the ALPACAS framework to address the design challenges of the collaborative anti-spam system.
Challenge 1. To protect e-mail privacy, it is obvious that the messages have to be encrypted. However, in order for the collaboration to be effective, the encryption mechanism has to satisfy two competing requirements: 1) The encryption mechanism has to
hide the actual contents for privacy protection and
2) It should retain important features of the message
Challenge 2. To avoid inference-based privacy breaches, it is necessary to minimize the information revealed during the collaboration process. However, the lesser the information conveyed, the harder it is to perform meaningful similarity comparisons.

Accordingly, the ALPACAS framework includes two unique components, namely feature-preserving fingerprint and privacy-preserving protocol to address the above challenges, respectively. In addition, in the interests of scalability, we design a DHT-based architecture for distributing ham/spam information among the collaborating entities. The ALPACAS framework essentially consists of a set of collaborative anti-spam agents. An e-mail agent can either be an entity that participates in the ALPACAS framework on behalf of an individual user, or it may represent an e-mail server having multiple users. Without loss of generality, in this paper, we assume that the e-mail agents represent individual users. Each e-mail agent of the ALPACAS framework maintains a spam knowledge base and a ham knowledge base, containing information about the known spam and ham e-mails. Fig. 1a shows the e-mail agent EA4 querying two other collaborative agents with partial information of an incoming message for the purpose of classification. Fig.b illustrates the internal mechanism of each e-mail agent: Upon receiving an e-mail, the respective e-mail agent transforms the message into a feature digest. It then uses part of the
feature digest to query a few other e-mail agents to check whether they have any information that could be used for classifying the e-mail.

D. Feature-Preserving Fingerprint

In our approach, the fingerprint of an e-mail is a set of digests that characterize the message content. The set of digests is referred to as the transformed feature set (TFSet) of the e-mail. The individual digests are called the feature elements (FEs). The TFSet of a message Ma is represented as TFSetðMaÞ. In the following sections, we will discuss how to generate TFSet and how to further enforce the privacy preservation.



$$TFset = [FE_1, FE_2, FE_3 \ldots FE_s]$$

Fig c Feature-preserving fingerprint technique

E. Privacy-PreservingCollaboration Protocol:

Feature-preserving fingerprint is just one level of privacy protection; the amount of information exchanged during collaboration can be further controlled for stronger privacy protection. In particular, we design the collaborative antispam system equipped with privacy-aware message exchange protocol based on the following spam/ham dichotomy that revealing the contents of a spam e-mail does not affect the privacy or confidentiality of the participants, whereas revealing information about a ham e-mail constitutes a privacy breach.

Our protocol works as follows: When an agent EAj receives a message Ma, EAj computes its TFSet : TFSetðMaÞ. It then sends a query message to other e-mail agents in the system to check whether they can provide any information related to Ma. However, instead of sending the entire TFSetðMaÞ as the query message to all agents, EAj sends a small subset of TFSetðMaÞ to a few other e-mail agents (the e-mail agents to which the query is sent is determined on the basis of the underlying structure. The subsets of TFSetðMaÞ included in the queries sent to various other e-mail agents need not be the same (our architecture optimizes the communication costs by sending nonoverlapping subsets to carefully chosen e-mail agents).

An e-mail agent that receives the query, say EAk, checks its spam and ham knowledge bases looking for entries that include the feature subset that it has received. A feature set is said to match a query message if the set contains all the Fes included in the query. Observe that there could be any number of entries in both spam and ham knowledge bases matching the partial feature set. For each matching entry in the spam knowledge base, EAk includes the complete TFSet of the entry in its response to EAj. However, for any matching ham entries, EAk sends back a small, randomly selected part of the TFSet. protocol. In this figure, the agent EA4 sends a query with the FE 815033 to EA7, which responds with a complete feature set of a matching spam e-mail and a partial feature set of a matching ham e-mail.

At the end of the collaboration protocol, EAj would have received information about any matching ham and spam e-mails (containing the feature set of the query) that have been received by other members in the collaborative group. For each matching spam e-mail, EAj receives its complete TFSet. For each matching ham e-mail, EAj receives a subset of its TFSet. EAj now computes the ratio
of MaxSpamOvlpðMaÞ to MaxHamOvlpðMaÞ and decides whether the Ma is spam or ham. MaxSpamOvlp is the maximum overlaps between the TFSet of the query message and the TFSets of all the matching spam e-mails, and MaxHamOvlp is similarly defined. In this paper, we use a simple classification strategy that is described as follows:
Score ¼.If the score is greater than a configurable threshold _, Ma is classified as spam.

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. If you are using A4-sized paper, please close this file and download the file for “MSW_A4_format”.

III. CONCLUSIONS

In this paper, we have presented the design and evaluation of ALPACAS, a privacy-aware collaborative spam filtering framework that provides strong privacy guarantees to the participating e-mail recipients. Our system has two novel features:
1) A feature-preserving transformation technique encodes the important characteristics of the e-mail into a set of hash values such that it is computationally impossible to reverse engineer the original e-mail.
2) A privacy preserving protocol enables the participating entities to share information about spam/ham messages while protecting them from inference-based privacy breaches. Our initial experiments show that the ALPACAS approach is
very effective in filtering spam, has high resilience toward various attacks, and provides strong privacy protection to the participating entities.

REFERENCES

[1] Z. Zhong, L. Ramaswamy, and K. Li, “Alpacas: A Large-Scale
Privacy-Aware Collaborative Anti-Spam System,” Proc. IEEE INFOCOM ’08, Apr. 2008.
[2] V. Schryver, Distributed Checksum Clearinghouse, http:// www.rhyolite.com/anti-spam/dcc/, Nov. 2005.
[3] Vipul’s Razor Anti-Spam System, Vipul Ved Prakash, http://

razor.sourceforge.net/, 2008.
[4] E.S. Raymond, Bogofilter: A Fast Open Source Bayesian Spam Filters,
http://bogofilter.sourceforge.net/, Nov. 2005.
[5] J. Jung and E. Sit, "An Empirical Study of Spam Traffic and
the Use of DNS Black Lists," Proc. Internet Measurement Conf.
(IMC '04), Oct. 2004.
[6] T. Meyer and B. Whateley, "SpamBayes: Effective Open-Source,
Bayesian Based, Email Classifications," Proc. First Email and Anti-SP(CEAS 04)AM conf