A Survey of Emerging Approaches to Spam Filtering

GODWIN CARUANA, Brunel University, U.K.
MAOZHEN LI, Brunel University, U.K. and Tongji University, China

From just an annoying characteristic of the electronic mail epoch, spam has evolved into an expensive resource and time-consuming problem. In this survey, we focus on emerging approaches to spam filtering built on recent developments in computing technologies. These include peer-to-peer computing, grid computing, semantic Web, and social networks. We also address a number of perspectives related to personalization and privacy in spam filtering. We conclude that, while important advancements have been made in spam filtering in recent years, high performance approaches remain to be explored due to the large scale of the problem.

Categories and Subject Descriptors: I.5.2 [**Pattern Recognition**]: Design Methodology—*Classifier design and evaluation*; I.5.4 [**Pattern Recognition**]: Applications—*Text processing*; I.2.6 [**Artificial Intelligence**]: Learning

General Terms: Algorithms, Performance

Additional Key Words and Phrases: Spam filtering, classifiers, architectures, distributed computing, peer-to-peer, grid, semantic

ACM Reference Format:
Caruana, G. and Li, M. 2012. A survey of emerging approaches to spam filtering. ACM Comput. Surv. 44, 2, Article 9 (February 2012), 27 pages.
DOI = 10.1145/2089125.2089129 http://doi.acm.org/10.1145/2089125.2089129

1. INTRODUCTION

Email has become one of the most ubiquitous modern day communication tools. It is estimated that 247 billion email messages were sent per day in 2009—a figure anticipated to double by 2013 [Radicati 2009]. The availability and accessibility of email has become a necessity for many. Its applications range from basic informal communication to a fully fledged and indispensable business platform. However, capitalizing on both its popularity as well as its ubiquity has created an opportunity for a lucrative business model based on unsolicited bulk email or rather spam, as it is more commonly referred to. The proliferation of spam has reached a considerable magnitude of *contaminates* the enabling communication infrastructures on a continuous basis.

Spam, or unsolicited bulk mail, varies in shape and form [Gomes et al. 2004]. Nonetheless, it tends to exhibit a number of similar traits in terms of structure, content, and diffusion approaches. There is a perceptible business reason and justification for its proliferation: from a spammer's perspective, the effort and cost of sending a

M. Li is also a visiting professor of the Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, China.
Authors' addresses: G. Caruana and M. Li, School of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, U.K.; emails: {Godwin.Caruana, Maozhen.Li}@brunel.ac.uk.

DOI 10.1145/2089125.2089129 http://doi.acm.org/10.1145/2089125.2089129

substantial number of emails is minimal, and the potential reach in terms of the magnitude of the available audience size is enormous [Paul et al. 2005]. The prospective profitability is clearly described by Gansterer et al. [2005], and the numbers speak for themselves—the overall cost of spam in 2009 was estimated at 130 billion U.S. dollars [Ferris Research 2008].

Spam filtering is intended to segregate ham from spam. Applied from a technical perspective (other approaches, exist, including regulatory and organizational, e.g.), filtering takes various shapes and forms. Common methods include email client extensions (filters) and filtering processes at the mail service-provider end. Simple Mail Transfer Protocol (SMTP) and machine learning-based approaches are popular implementations of filtering schemes [Blanzieri and Bryl 2008]. Broadly speaking, SMTP-based methods generally refer to scrutiny of SMTP traffic, email exchange route verification, and authenticated SMTP sessions. The application of heuristics, based around techniques such as *black/white listing* (the application of trusted and non-trusted sources) [Levine 2005] are also common. Machine-learning techniques involve the analysis of email, mostly at content level, and employ classification algorithms, such as Bayesian, Support Vector Machines, and others to segregate spam from legitimate email. These approaches have been extensively applied in spam filtering and exhibit different capabilities [Hunt and Carpinter 2006]. Schyryen [2007] states that combinations of techniques and algorithms employed together, or *cocktails* (as more creatively referred to in Conry-Murray [2004]), significantly increase the potential to alleviate spam-related issues. Given the scale of the spam problem, unless careful mitigating controls and filtering considerations are applied from the outset, the value proposition of email and associated services may degrade rapidly overtime. This degradation is contextualized in terms of overall usability and value for money, as spam starts to deeply permeate respective email infrastructures.

Although there have been earlier attempts to review spam-filtering methods [Guzella and Caminhas 2009; Blanzieri and Bryl 2008], the scope is mainly related to machine-learning approaches, which are usually deployed on small-scale computing environments targeting low volumes of spam messages. The proliferation of spam demands high-performance-filtering algorithms and methods.

The past few years have witnessed a number of emerging technologies in computing, notably, peer-to-peer computing [Schollmeier 2002], grid computing [Li and Baker 2005], semantic web [Lee et al. 2001], and social networks [Boyd and Ellison 2007]. These technologies facilitate the development of collaborative computing environments for solving large-scale problems. A number of approaches to spam filtering have been proposed and implemented, building on these emerging computing technologies. This survey is targeted at emerging approaches to spam filtering which complement existing surveys. Additionally, security- and privacy-related issues in spam filtering are further discussed.

The remainder of the article is organized as follows. Section 2 provides an overview on traditional approaches to spam filtering. Section 3 reviews emerging approaches to spam filtering and discusses the challenges that these approaches present. Section 4 discusses various issues related to security and privacy in spam filtering. Section 5 presents a number of observations identified in this survey and discusses research directions. Section 6 concludes the article.

2. AN OVERVIEW OF TRADITIONAL APPROACHES TO SPAM FILTERING

As mentioned in Section 1, SMTP and machine learning are popular implementation approaches to spam filtering. An overview of these is provided, accompanied by a discussion of a number of core challenges associated with them.

2.1. SMTP Approaches

The SMTP protocol is at the foundation of the email enabling infrastructure. It is regularly argued that a number of the limitations capitalized on by spammers are inherent to its original design [Roman et al. 2005] since one of the Primary concerns was to preserve. The motivator behind the work presented by Duan et al. [2007], for example, is primarily focused on addressing a number of these simplicity limitations. The basic approach presented tries to shift the control from the sender to the receiver, in order to mitigate the unwanted reception of unsolicited mail. This is achieved by changing, in part, the delivery approach from *sender-push* to *receiver-pull* [Duan et al. 2007]. The proposed extension to the baseline protocol, designated Differentiated Mail Transfer Protocol (DMTP), introduces an additional measure for email transactions. The recipient is first sent an *intent* message [Duan et al. 2007], subsequent to which the message content is only retrieved if interest in the former is shown. While this approach exhibits substantial potential, any change to the underlying core email-transmission infrastructure is bound to introduce operational complexities. Furthermore, it is prone potentially to impact a number of intertwined enabling services, even if an incremental approach to its deployment is undertaken, as suggested. This risk has the potential to considerably adoption and acceptance levels for such an approach.

SMTP protocol inherent at its core is also evident from the very small set of operations allowed, namely HELO, *MAIL FROM*, *RCPT TO*, *DATA*, . , *QUIT*, and *RSET*, in that order. To this extent, Li and Mu [2009] base their spam-filtering approach by scrutinizing SMTP traffic payload data. More specifically, abnormal SMTP interaction commands are used as rationale, such that unusual application of SMTP traffic is intended to enable high throughput. It is argued that these approaches are not representative of normal email users or email clients' interaction patterns. This rationale is subsequently employed as the basis for assuming that such traffic is most likely spam. Designated Abnormal SMTP Command Identification (ASCI), the results from the work presented by Li and Mu [2009], indicates that this approach can achieve around 11% reduction in spam. The endorsement of normal protocol interaction by spammers or bots will undoubtedly hinder the spam-sending throughput rate considerably. However, adopted as is, ASCI runs the risk of not retaining original effectiveness over time. This challenge manifests itself when intelligent botnets are involved and able to identify and react to basic nondelivery issues. Irrespectively, this approach provides a representative first level of defense in reducing the number of spam that actually infiltrates inwards from the edge network.

Another prevalent technique for filtering spam is *Greylisting*. While numerous variations exist, the core concept is rather simple. The IP address of the SMTP host participating in the exchange and the addresses of both the sender and recipients (referred to as a *triplet*) are employed to *deny* the email exchange of happening initially. Triplet information is stored, and subsequent attempts with the same signature are allowed to perform the mail exchange. This approach is based on the rationale that most spammers do not care to retry sending the same email from the same relay. Numerous studies supported by empirical results [Levine 2005; Twining et al. 2004; Gansterer and Ilger 2007] show that this approach is rather effective. Furthermore, it is nonintrusive from an enduser perspective, which is, considered a critical success factor. The combination of Greylisting with other techniques, such as Blacklisting, further increases the potential to control spam. Blacklisting, or real-time blacklist (RBL), involves setting up and maintaining a list or lists of sources, including open relays, that are identified as spam propagation sources. Numerous sources for Blacklists and associated services exist, notably Spamhaus [Spamhaus 2009], SpamCop [SpamCop 2009], and DNSBL [DNSBL 2009]. Innovative variations of Blacklisting-based approaches exist as well [Ramachandran et al. 2007].

Furthermore, non SMTP approaches also exist can be classified as infrastructure-related. Spam can be stopped at the source before it leaves its origin, in transit, or at the receiving end. Various degrees of filtering and combination points exist [Goodman and Roundthwaite 2004]. Approaches include those which sit beyond the individual point of spam origin but still within the source network. Other inspection and filtering points sit at the very edge of the destination email service-provider network. Gateway-and-router based approaches are popular, as well. One such example is presented and discussed in Cook et al. [2006]. One can refer to this approach as *late pre send* given that filtering happens just before the email is dispatched out of the originating source network. Intrusion detection systems (IDS) are regularly employed near the entry point of enterprise networks. The method presented in Cook et al. [2006] capitalizes on this approach, albeit this method could also be employed at different parts of the network. By inspecting various information sources that the IDS gathers, assembles, and has access to, as well as the ability to identify correlations, the approach performs dynamic, domain-specific blacklisting. IP address domains which are identified as spam sources are blocked via the automated insertion of respective firewall blocking rules, based on the intelligence gathered and built by the environment. This type of approach avoids spam flowing into the network in the first place. However, the effectiveness of this approach depends on a number of sources which are not fully under the control of the solution itself. This may severely hinder its overall effectiveness in a high throughput production system.

2.2. Machine-Learning Techniques

A number of filtering algorithms are employed from a technology standpoint, varying in terms of complexity and effectiveness. However, spam filtering algorithms can be split into two overall umbrella approaches, namely *machine* and *non-machine-learning* methods. Approaches applied in the former category include Bayesian classification, neural-networks, Markov-based models, and pattern discovery. Rule, signature- and hash-based identification, blacklisting, and traffic analysis, among others, are techniques that are employed with respect to non-machine-learning variants. Both classes have their advantages and disadvantages and demand different levels of requirements, such as processing and bandwidth.

Machine-learning variants can normally achieve effectiveness with less manual intervention and are more adaptive to continued changes in spam patterns. Furthermore, they do not depend on any predefined rulesets analogous with non-machine-learning counterparts. Two principal machine-learning approaches for inferring spam classification in a semi- or fully autonomous fashion are commonly considered, namely supervised and unsupervised. The former depends on an initial training set to assert classification, while the latter does not, employing rather other techniques, such as clustering, to achieve its objectives. In supervised learning, model input observations, more formally referred to as *labels*, are associated with corresponding outputs upfront; in nonsupervised approaches, any observations are associated with latent or inferred variables.

From a high-level perspective, considering spam from a text-classification dimension in the context of machine learning, one can describe the basic classification problem via the dichotomy $\mathrm{m}_i \in \{-1, +1\}$. Here m_i represents a set of messages and -1 and $+1$ represent nonspam and spam respectively. In order to apply text classification to spam, messages are normally represented as a set of vectors, such as $\{\} : \mathrm{m} \rightarrow \mathbf{R}^{n.}$ Here the vector set $\mathbf{R}$ represents the features of a message. Each message has a corresponding feature vector constructed using features. Identifying those features of a message which have the qualities to indicate whether it is spam is a critical task for machine-learning approaches. This also includes, where applicable, a number of preprocessing tasks, such

as lexical analysis and dimensionality reduction, in the form of stemming, cleansing, and normalization [Sebastiani 2002]. Features can take the form of words, combinations of words and phrases, etc. Generally speaking, fewer features normally represent greater generalization and better performance, however, mostly at the expense of not being able to obtain the required class separation, that is the identification of the optimal separating level between the classes (spam/ham), thus minimizing or removing entropy. Various schemes intended to identify the best features in terms of quality and number are possible, as discussed by [Yang and Pedersen 1997]. As indicated earlier, feature vectors are typically associated with weights intended to influence the outcome of the classification accordingly. Popular weighting schemes include term frequency (TF), binary representation, and term frequency–inverse document frequency (TF-IDF) [Salton and Buckley 1998].

Besides the definition of the respective feature vectors and subsequent creation of the training set, there is also the classification algorithm itself that needs to be considered. As indicated earlier, numerous machine-learning algorithms exist, including Decision Trees, Bayesian classifiers, k Nearest Neighbor (kNN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM). While the specific review and discussion of machine learning-based approaches and their accompanying algorithms is not the scope of this work, these approaches continue to garner a lot of attention in the context of spam filtering [Guzella and Caminhas 2009; Blanzieri and Bryl 2008; Hidalgo 2005].

2.3. General Challenges

Although significant advancements in spam filtering have been made with traditional approaches, a number of challenges remain [Goodman et al. 2007]. SMTP is at the core of the email-enabling infrastructure, so it follows that considerable effort to handle spam at this level has been applied, including research from various perspectives [Duan et al. 2007; Gburzynski and Maitan 2004; Bernstein 2000]. The biggest challenge for SMTP-based approaches is to ensure that they do not impact in any way the underlying and enabling infrastructure. The number of components and building blocks involved in successful email exchange and which are directly dependent on the protocol itself are not trivial. While the potential is there, technical complexities as well as financial challenges restrict (in various ways and to different extents) the widespread consideration and adoption of SMTP-based approaches. As an example, email exchange-path verification, which provides the ability to trace the real origin of an email sender, requires appropriate accounting in the context of the packet network involved. Authentication, on the other hand, requires appropriate client support as well as cooperation, and therefore may limit the clients that are able to interact in such fashion. While the popularity of Blacklisting and Whitelisting continues [Blanzieri and Bryl 2008; Goodman et al. 2007; Janecek et al. 2008; Jung and Sit 2004; Khanal et al. 2007] challenges still exist, including the increased possibility of blocking legitimate email exchange, or false positives, which is considered to be very costly. Manual interventions for applying changes or adding new records to these lists make them prone to mistakes. There is also an additional burden of keeping them up-to-date. Similarly, where heuristic-based approaches are employed, it is difficult to maintain the necessary patterns that are employed for matching up-to-date.

Different challenges also exist in machine-learning approaches to spam filtering [Goodman et al. 2007]. Specific algorithms have particular advantages and disadvantages which influence their overall accuracy and performance. This has been extensively discussed in numerous related works which compare algorithms and approaches from a research [Zhang et al. 2004; Kolz and Yih 2007; Sebastiani 2002], as well as real-world application perspectives [Cormack and Lynam 2007; Hayati and Potdar 2008]. From an end user standpoint, what the email user perceives as spam and not

spam remains an active research question, due to the degree of personal subjectivity associated with its classification Spam or not Spam—That is the Question [Kiran and Atmosukarto 2009]). Challenges surrounding false positives, including the implications they can lead to, have also been discussed extensively. False positives refer to wrong outcomes from classification schemes. More specifically, they portray the situation whereby ham is classified as spam. While such challenges can be somewhat mitigated by the concurrent application of a number of different classification schemes (classifier combining) as discussed in Hershkop and Stolfo [2005], they still constitute an ongoing challenge. The same applies to security and other typical challenges associated with machine learning [Barreno et al. 2006].

The overall utility of a classifier directly depends on the training set [Weiss and Tian 2006]. The training element of many machine-learning approaches requires content from real-world email in order to be as representative as possible. Quality training data may not be readily available or tailored to an extent that may impact the overall quality of the learning models. This challenge is amplified further when the learning data is frequently changing and mandates continued efforts in terms of retraining. The performance of numerous machine-learning approaches is also largely dependent on the size of the training set. Although a larger training set would produce better results, more processing time in terms of the learning process is normally required. Furthermore, the identification and selection of the best feature set introduces further challenges. Learning data tends to reflect specific attributes, perhaps not distinguishing enough between the relative weights of attributes, as well as missing relevant properties. Changes in spammer behavior, as well as the delivery type of spam payload, have also introduced different sets of challenges which require innovative approaches to remain effective in spam filtering.

Additionally, the traditional Mail User Agent and Mail Transfer Agent model for spam filtering faces continued challenges related to scalability and performance. The sheer number, type, and size of spam traversing communication exchange paths mandates considerable computing resource requirements for filtering. These requirements tend to be of a magnitude beyond most traditional filtering architectures. It is extremely difficult for traditional spam-filtering architectures to be in position to scale up (and down) at the rate mandated by the fluctuations of spam proliferation.

Summing up, traditional approaches to spam filtering face continued and increasing challenges. These approaches are usually deployed on computer nodes which process spam individually without collaboration, limiting their application to a small scale.

3. EMERGING APPROACHES

Spam and spammer techniques evolve through time, capitalizing on new approaches and exploiting new flaws. Building on recent development in of computing technologies—notably peer-to-peer (P2P) computing, grid computing, semantic web, and social networks—a number of emerging approaches have been proposed for spam filtering. These approaches are intended to tackle a number of challenges briefly discussed in Section 2. They are also intended to improve overall spam filtering effectiveness. The following sections discuss representative research work in these areas.

3.1. Peer-to-Peer Computing

The ability to capitalize on distributed resources, including hardware, software, as well as human participation is what constitutes and drives the core of P2P initiatives and architectures (see Figure 1). From a very high level perspective, there are two *umbrella* types of peer to peer (P2P) architectures, namely centralized and decentralized. Decentralized P2P networks can be further classified into unstructured and structured P2P networks.

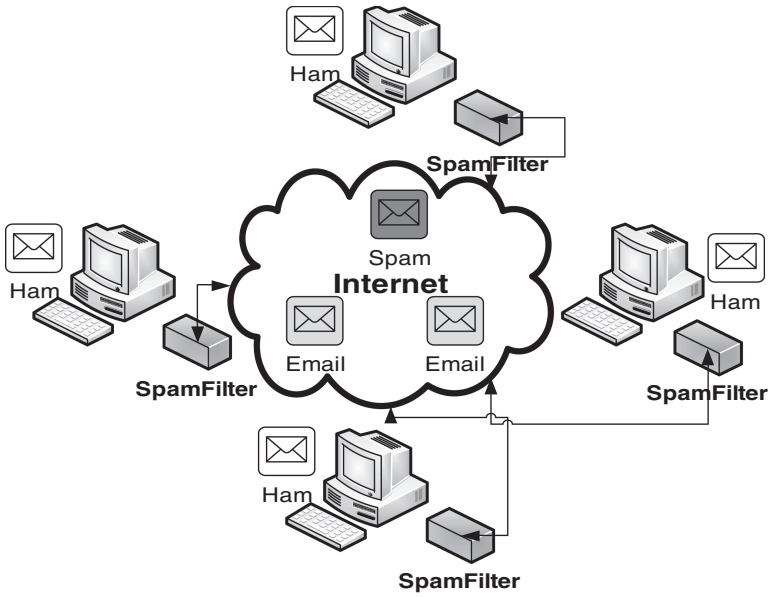


Fig. 1. Peer-to-peer computing for spam filtering.

P2P architectures form the backbone of numerous distributed, high-performance services and systems, significantly popularized in recent years by Internet-based application-sharing communities and software. These include those based on the Gnutella [Gnutella 2001] and FastTrack protocol (undocumented formally), such as Limewire [Limewire 2009]. The ability to exploit high scalability easily, as well as offer a good degree of personalization, are inherent potentials of P2P-based systems. Furthermore, the ability to make efficient use of common interests is very inductive to collaboration. In this context, P2P computing application for spam filtering is applied in a number of ways. Peers can collaborate in the identification and filtering of spam via the exchange of either computing power or, more simply, intelligence gathered on email and spam. Various techniques can be employed; however the algorithms and techniques employed in this context are normally focused on minimizing network bandwidth and identification of nearest peers. Exchange of information is normally focused on precomputed *signatures* which relate to previously scrutinized and tagged content. The structuring element plays a critical role in P2P systems. Flat, hierarchical, and distributed hash tables (DHTs) are commonly considered. Specifically with respect to DHT-structured P2P networks, popular implementations such as Pastry [Rowstron and Drushel 2001] and Chord [Stoica et al. 2001] provide sound architectures that are scalable and are inherently fault-tolerant. Dimmock and Maddison [2004] provide a review on the application of P2P systems to spam filtering in this context.

The CASSANDRA (Collaborative Anti-Spam System Allowing Node-Decentralized Research Algorithms) initiative presents a personalized approach for tagging and classifying spam [Gray and Haahr 2004]. The work employs a P2P architecture and exploits computing resources that participating nodes can offer to achieve high scalability and flexibility. The authors argue that, while common spam filtering established on traditional nonmachine- or machine-learning approaches and based on mail content are reasonably effective, they are still considerably prone to false positives. Collaborative spam filtering tends to be better suited to tackling issues related to spam *drift*, or rather, spam and spammers tendency metamorphoses in trying to circumnavigate

filters based solely on content. The prototype architecture described, however, does not seem to consider possible issues related to how it can distinguish between real-world users and automated spam bots.

Building on a Chord P2P network, Dimmock and Maddison [2004] present another approach to exchanging partial and hashed message data on spam between participating nodes. This allows the network to distinguish good participation from malevolent. However, one challenging issue is that spammers could easily influence the P2P network by changing their participating status from nontrusted to trusted participating entities. In the real world, this may cause the approach to gradually loose efficiency over time. Being a proxy-based implementation allows for great flexibility in terms of the usability from the widest variety of email clients (assuming adequate levels of heterogeneity). However, it also introduces an additional element of complexity, as well as computational resource requirements at the client end. This challenge may be further amplified by the fact that the proxy itself is also a simple Web server, increasing the local node's attack surface from a botnet perspective. Updates to the proxy itself can also prove difficult given the potentially high distribution of nodes which will be making use of the localized proxy, Other P2P-based approaches to spam filtering include those presented by Luo et al. [2007], Damiani et al. [2004], Zhou et al. [2003], and Metzger et al. [2002]. Foukia et al. [2006] discuss a diverse approach. The primary motivation of their work concerns the collaboration of email servers, rather than the collaboration between mail clients. Spam checking is primarily performed at two specific points in time: before being sent by the sending email service and at the receiving email service end. A system of rewards and restrictions is applied to participating servers. The overall standing of these servers with regards to *trust* a directly effects the degree of influence on the resolution. A critical factor for the effectiveness of this approach is the participation of a considerable number of email service providers (ESPs). Effectiveness is drastically decreased if participation is low. Additionally, the possibility for legitimate email to be slowed down due to the respective ESP being classified as a temporary spam source could potentially irritate legitimate email users. The work does, however, indicate possible avenues for exploration. The discussed approach does not consider the utilization of resources available at the client's end for additional collaboration and contribution towards the improvement of spam detection and filtering. It can therefore be argued that this reduces possible input sources, as well as overlooks the possibility of benefitting from additional processing power for improving intelligence.

A rather distinctive viewpoint using Percolation Networks as a theoretical base is provided by Kong et al. [2006], client-side plugins in conjunction with, which uses common email clients. This approach identifies spam using a typical classification scheme (such as Bayes). A digest function, which the client must be able to generate, is subsequently produced if the email is identified as spam. The client then publishes this newly identified intelligence to a number of participating *neighbors*. This is done using a *random walk* path through the percolation network based on the graph edges (*distance)*. Mail clients can query the network to identify whether an email has already been tagged as spam. This scheme works by implanting the inquiry using the same approach employed for publishing its spam digest database updates, that is, by traversing the percolation graph using a random walk to a prespecified distance. The query percolates and aggregates digest hits across nodes, returning the results of the query via a reverse path. If the hit count meets a specified score, the message on which the query digest was based is declared as spam. Data exchange during querying and publishing does not appear to be high, but a degradation of performance from an underlying network perspective still has the potential to inflict nontrivial performance penalties, reducing the overall effectiveness of this approach.

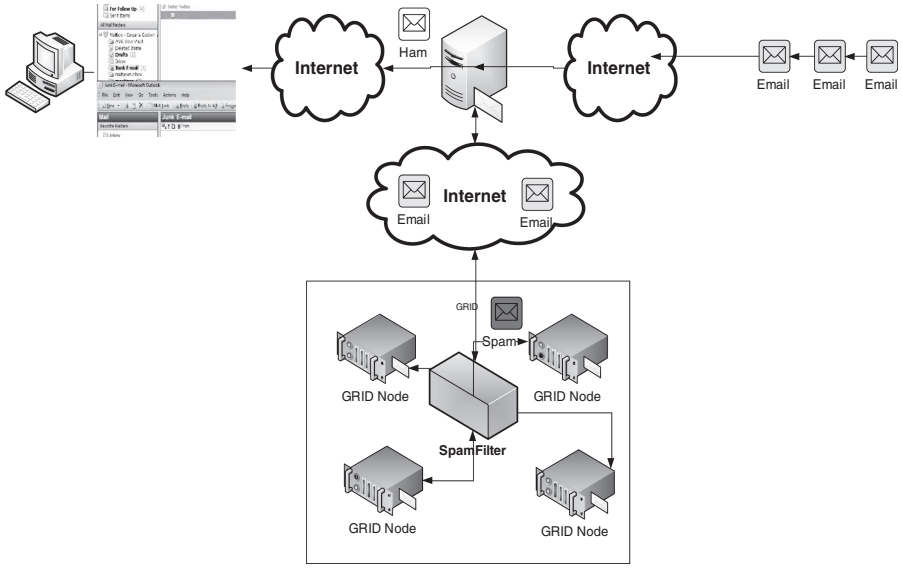


Fig. 2. Grid computing for spam filtering.

Overall, in the context of spam filtering, P2P approaches provide a number of advantages by exploiting the underlying mesh-based network infrastructure. P2P networks are well suited for collaborative spam filtering. The computing capacities of P2P networks can easily grow with the increasing number of peer nodes. Ad hoc distribution of messages also eliminates single point of failures and permits active distributed processing. Another opportunity which can be exploited is aggregation of participants with similar interests in spam filtering. However, P2P networks suffer from a number of issues: unstructured P2P networks are technically challenged in achieving high scalability in locating peers. They are also prone to a number of security challenges. In index poisoning, for example, when a resource search is performed, the results returned may not reflect correct information because bogus information may have been inserted in the respective tracking structures [Liang et al. 2005]. Also, in structured P2P networks, while DHT approaches provide the necessary scalability, they suffer from complex and challenging overheads associated with structure maintenance, as well as management [Kurose and Ross 2004].

3.2. Grid Computing

Originally conceptualized in 1997, grid computing has evolved into an effective computing paradigm for solving data and computationally intensive problems by utilizing various computing resources over the Internet [Li and Baker 2005]. Figure 2 portrays a general grid-computing architecture for spam filtering.

Grid computing promotes the utilization of collective computing resources that facilitate spam filtering on a large scale. Rather than investing in resources to keep up with continued spam increase, capitalizing on grid resources on a need basis has significant advantages. Hosted and Cloud-based anti-spam service implementations such as Microsoft [2010] and GFI [2010], in which the spam filtering element is contracted out to a third party, are also commonly based on grid computing principles. Similar to P2P computing, the ability to share resources, as well as intelligence, in a collaborative fashion provides further opportunities improving the quality of spam filtering.

Liu et al. [2005] present a grid-based distributed-Bayesian classification scheme for spam identification in a collaborative fashion. Specific algorithms are employed to reduce the number of false positives which are inherent to common Bayesian approaches. The authors argue that the grid approach is more efficient than P2P-based counterparts. The key challenge with this approach is the dependency on specific plugins for specific email clients, which is, arguably, architecturally brittle (due to client stack and version differences), as well as tightly coupled, which limits its flexibility. The notion of a virtual organization in the context of grid architectures, which creates the opportunity for a worldwide anti-spam organization, is mentioned in this article. In such a scenario, the virtual organization can offer a number of associated services ranging from providing computing resources for spam filtering to legal frameworks for spam control. However, limited discussion regarding this is articulated. Another grid-based architecture initiative is discussed at a very high level in Yuan et al. [2007] which utilizes a global grid infrastructure for gathering spam intelligence.

In principle, grid-based approaches benefit from unlimited scalability potential. The same applies to the ability to parallelize jobs in the context of machine learning for large-scale spam filtering. The ability to use computing resources on a need basis is also considered an important attribute, due to limited capacity in planning visibility with respect to spam increase, decrease, and fluctuations. Furthermore, underutilized resources can be capitalized upon for high-volume spam filtering, thus making effective use of idle resources. However, a grid is a large-scale computing environment in nature, and grid resources are highly heterogeneous, thus carrying an obvious element of complexity. A number of additional challenges remain in grid-based approaches to spam filtering, including the discovery of and load-balancing among resources, and that standards of grid computing have not yet reached maturity [Li and Baker 2005].

3.3. Social Networks

Social networks are arguably one of the most remarkable Internet phenomena in recent years. They have emerged as one of the most-sought applications of the World Wide Web—a *killer app*. They provide a ubiquitous ecosystem which allows users to identify themselves, interact, share, and collaborate. Alongside similar work by applied Ohfuku and Matsuura [2006] and Boykin and Roychowdhury [2005], Chirita et al. [2005] pose a research question in the context of spam using social networks for gathering intelligence. Analogous in a number of ways to PageRank [Brin and Page 1998], MailRank is introduced in this study as a ranking and classification scheme. Two approaches are described: one basic, the other offering personalization. In both cases, the primary technique is the aggregation of email intelligence from participants of the social network. A network graph is constructed relating scores to each respective email address. By employing a power iteration algorithm, a reputation-weighting measure which is influenced by participating members via the exchange of trust votes is associated with emails. This architecture relies on the introduction of the centralized service designated MailRank, as well as an intermediary proxy intended to sit between the Mail User Agent and the production email service.

Another application of social networks in the context of spam filtering is discussed in Garg et al. [2006]. The primary motivator here is the exchange of filters between respective social network participants, rather than exchanging signatures or similar statistics to increase overall spam intelligence. The authors argue that there are considerable benefits from capitalizing on end-user processing capabilities in contrast to purely centralized approaches. This work also introduces the notion of a spam filter description language and a pluggable engine able to make use of filters which adhere to the described specification. This creates a way to describe spam filters and associated

behavior in a uniform way, which can be subsequently exploited by architectures which are aware of its specification.

Hayati and Potdar evaluate Web 2.0-based anti-spam methods [2009]. Their result highlights the need for more sophisticated measures for combating spam in the social networking sphere, while ensuring that any applied measure remains as nonintrusive as possible. Zinman and Donath also present an innovative approach for tackling spam in the context of social networking [2007]. The focus and motive of this work is the ability to categorize participants according to their *characters* using the concept of *feature bundles*. The conceptual features describe prototypes, which include events, actions, emoticons, spatial relationships, and social relationships. The final objective is to categorize users via two primary dimensions, namely sociability and influence. The former is employed as a metric describing the level and nature of information; the later is based on the influencing element that the profile tries to carry. The approach employs combinations of these metrics in turn to identify the type of user. For example, a user with low sociability and high promotion is initially evaluated as an entity whose primary intent within the network is marketing and expansion.

An interesting approach for specialized social networks dealing with videos is discussed in Benevenuto et al. [2008]. The challenge is identifying whether video content is ham or spam. The authors employ heuristics from a machine-learning approach based on video attributes in conjunction with a user's behavior to perform classification. Characterization is based on comments submitted and popularity in terms of number of views. Video spam is denoted if, for example, comments submitted are not related to the video description or have zero-length content. Here, the challenge of what is considered spam to whom emerges; that is, the degree of subjectivity of what *is* spam is even greater than textual counterparts because of increased difficulty and complexity in characterizing video content.

Similar to P2P, social network-based approaches yield a number of potential intelligence-collection sources. Social networks also tend to aggregate participants with similar interests, promoting collaborative environments which benefit spam filtering. Controlling, and filtering spam in a social network context is generally more difficult than compared to traditional email exchange ecosystems [Brown et al. 2008]. Given the complexity in identifying and segregating participants' intents, it is also prone to generic poisoning attacks [Yeh et al. 2007]. Similar to P2P, these poisoning attacks refer to the intentional introduction of influencing factors intended to bias outcomes towards specific objectives which relate to the environment.

3.4. Ontology-Based Semantics

Ontologies have played a critical role in the arena of the Semantic Web. They provide a formal basis for the definition of knowledge representation, subsequently enabling the exchange of this knowledge relatively easily. Ontologies are used to describe specialized domains and an associated set of vocabularies. Semantics relate to the ability to portray and understand the meaning of information in an expressive way. Their combination in the context of spam filtering facilitates the definition and understanding of spam in a better and formal way. The ability to exchange intelligence and, subsequently, the potential to process it in a formal and interoperable fashion provides numerous advantages. In the context of spam, ontologies are employed to represent email content and spam, as well as users' preferences in terms of spam definition, ensuring a consistent interpretation of filtering schemes.

To this extent, the work articulated in Youn and McLeod [2009a] argues that the application of ontologies to formalize spam offers a sound basis and numerous opportunities for improving the definition and filtering of spam, in the context of reflecting user preferences more appropriately. Putting greater emphasis on the personalization

aspect will greatly increase the perceived usability. Ontological knowledge is built by identifying and formalizing the relationship between a user choices and how spam is reacted to, that is presented, for example, in the work by Kim et al. [2007]. The authors classify *reaction* into four types: *Reply*, *Delete*, *Store*, and *Spam*. On this basis, association mining is applied to an initial reference dataset using Weka [Witten and Eibe 2005]. Although the quality of the outcome (in terms of correlating user preference with typical reactions to specific emails) was constrained by the limited initial scope and entropy of the data set, the outcome still indicates the potential of ontologies in spam filtering. This approach could potentialy offer greater scope in terms of classification than conventional spam classification schemes, due to the real-world connection that exists between the email recipients and the way in which received email is tackled and actioned. Rather than classifying an email as spam solely based on typical machine-learning content-based techniques, the actual action of the recipient is taken into consideration. This approach influences the classification by differentiation of the action performed, as well.

Gray mail is a class of email that does not exhibit sufficient traits for establishing a degree of confidence that it is either spam or ham. This challenge is amplified further from a personalization perspective, given the associated subjectivity. These sets of attributes make the study of ontology-based approaches a motivating consideration in the context of gray mail. Youn and McLeod [2009b] present a spam filter based on a personalized ontology. Yin et al. [2007] present a different approach. They emphasize the advantages of a multipronged approach, comprising globally trained datasets for generalization and personalized equivalents for specialization.

Hsia and Chen describe another approach for image-based spam detection [2009]. In this work, a scheme based on exploiting hidden topics within images is employed. These latent topics are identified (and subsequently employed) as training input for a binary classifier. The authors describe a probabilistic approach for inferring hidden semantic meanings represented as images. In Youn and McLeod [2009b], the image element of spam representation is tackled using a traditional approach based on optical character recognition, and term frequency–inverse document frequency (TF-IDF) feature set selection is applied. Additional processing is performed to convert the model generated via the adopted machine-learning scheme, using Weka to RDF (Resource Description Framework). This step is employed in order to generate the required ontologies that are subsequently employed to create custom user filters.

The annotation of spam and email messages with metadata has benefits, including augmented intelligence, context richness, and formalization. The incorporation of domain knowledge facilitates filtering processes, including training and classification for high accuracy in spam filtering. Furthermore, ontologies bridge the gap between the levels of understanding required for preparing training and classification models and the end user. Due to the inherent readability and expressiveness elements, ontologies provide end-users with the opportunity to understand and contribute improving spam filtering which exploits such approaches. To date, however, there is no standard ontology for spam annotation, furthermore, separate initiatives tend to develop distinctive ontologies, this creating challenging situations with respect to ontological interoperability requirements. Ontology-based interoperability is by no means not trivial, [Orgun et al. 2006] as it becomes further amplified given the subjectivity aspects. Unless interoperation is possible, the benefit of sharing ontologies is lost, resulting in duplication of efforts.

3.5. Other Approaches

Spammer behavior study is a very important source for gaining wider insight from an overall spam-detection and filtering perspective, and may be supported by providing

ecosystems that allow spammer roaming while ensuring appropriate levels of monitoring. Gathering as much information as possible on how spammers operate is considered fundamental to identifying mitigating factors [Ramachandran and Feamster 2006; Pathak et al. 2008; Calais et al. 2009]. Spammers tend to rely on aggregated intelligence to work out target email addresses. "A taste of ones medicine" can be used as an idiomatic narrative of the work presented in Antonopoulos et al. [2009]. Albeit perhaps questionable from an ethical perspective, the work presented considers an original perspective, in an attempt to revert some of a advantages of a spammer's business model. The ultimate objective is to make spam more expensive or comparably less effective overall than alternate marketing schemes. The authors advocate that this will make spammers abandon their preferred activity of making money. One of the most important assets spammers have access to is a list (referred to as a database in the text) of email addresses, which defines the ultimate spamming audience. The work discusses the proposition of poisoning such a list in a collaborative fashion. The authors assert that, to date, the validity of these email addresses is close to 100%, and that a substantial reduction in quality would decrease the overall profitability of the spammer's business model. The suggested modus operandi is rather simple—inject routable fake email address to the database to inflate it with bogus records. They also suggest respective anti-spam resources increasing so that these newly acquired resources may be employed simply to mimic spam interaction, while, in fact, spam content is simply discarded rather than processed. This should result in less spam being received and processed by real end-users, assuming no change in the amount of spam being sent. Shinjo et al. present an approach in which the filtering elements are considered as features [2010] that can be enabled and disabled according to a set of attributes. These are referred to as capabilities exhibited by the email in transit. Capabilities are generated by a recipient and submitted to authorized senders from the context of the capability creator. If the email exhibits valid capabilities, the spam filter simply disengages its action and allows the email to be passed directly to the recipient's mailbox. One of the primary motivators behind this scheme is to eliminate false positive challenges. Because the implementation of capabilities is based on a standard approach using special SMTP headers, Mail Transfer Agent (MTA) and Mail User Agents (MUA) are able to handle or ignore the special directive(s) out of the box; however, extensions to the MUA are required to process the capabilities.

Esquivel et al. present another approach to limiting the proliferation of spam from its sources [2009] based on passive TCP fingerprinting at the router level. Here, SMTP servers collaborate with routers by computing signatures that represent spam sources, thus propagating periodical updates. Fingerprinting capabilities are required on both the SMTP servers, as well as the participating routers. The risk of overloading the high-throughput resources (such as routers) dues exist, and furthermore, the degree of intelligence that can be applied at this point to ensure respectable throughput, including ensuring zero false positives, is limited.

4. SECURITY AND PRIVACY CONSIDERATIONS

Email is considered mostly a means of personal communication in the context of participants' selectiveness. Therefore, security and privacy aspects are fundamentally important.

4.1. Security Aspects

An efficient way of hiding a spammer's identity and the spam's source of origin is by hijacking computing resources. These resources, which do not belong to the spammer, are subsequently employed to perform spamming operations, and thus, appear as spam sources themselves. This can be done either on a transient basis or for longer

periods without the legitimate resource owner's knowledge. Although some sources report a slow down in botnet activity [Secureworks 2009], botnets remain a real nuisance when dealing with spam proliferation due to the continued increase in hijacking sophistication, distribution, and execution capabilities. Various studies on botnets have been conducted, researching their behaviors, characteristics, and methods of dealing with them. Brodsky and Brodsky [2007] provide an interesting perspective in describing Trinity. Based on P2P-architecture and implemented as a *SpamAssassin* plugin [SpamAssassin 2009], Trinity is founded on the premise that automated botnets send a large amount of unsolicited email in a relatively short period of time. Participating peers process and provide information related to the mail relays they are associated with and have information about. This exchange of information is used to measure associated email-sending rates. Another detailed study on characterizing botnets, provided by Xie et al. [2008], presents a framework designated *AutoRe*, which is intended to provide a signature generation framework for identifying botnet-sourced spam. Similar work on a malware-generated email-identification method using clustering algorithms (based on a modified Levenshtein distance scheme and combined with Jaccard similarity coefficients) is described in Wei et al. [2009].

To mitigate source counterfeiting and introduce additional traceability elements from the sender's end, email authentication [Allman 2006] has been researches and on the application map [Lawton 2005; Peterson 2006]. Numerous efforts to this extent can be identified—notably, the Sender ID Framework (SIDF) and the Domain Keys Identified Email (DKIE)—however, they do not seem to have reached the critical mass as originally expected. Lawton [2005] argues that large enterprises with substantial numbers of email addresses may face challenges when trying to adopt these approaches. In addition to the concern of potentially breaking typical mail-forwarding capabilities, SIDF is also prone to having spammers registering as legitimate users. Lawton articulates that in 2005, a noticeable percentage of emails which had SIDF records originated from domains normally associated with spammers.

DKIE is heavily based on Public Key Cryptography, according to Liao and Schwenk [2007] who argue how Secure/Multipurpose Internet Mail Extensions (S/MIME), which employed by DKIE, can be improved. They suggest that ensuring end-to-end integrity from sender to recipient, while retaining full backward compatibility with the original implementation. Ironically, some argue that spammers themselves are among the more avid adopters of this approach [The Register 2004]—where authentication is applied in isolation, spammers endorsing this technique have a golden key for sending unsolicited bulk email.

Another presend filtering technique intended to stop email at the source rather than trying to control it at the recipient end is presented by Wei et al. [2009]. The basic concept is based on the ability to ensure the legitimacy of the source. Email that is not verified using an e-stamp is forwarded for spam scrutiny. Here, the e-stamping scheme is based on IAPP, an integrated authentication process platform and it is employed in conjunction with dynamic black or white listing, behavioral identification, and Bayesian techniques. Conceptually, the originator of the email marks the source with an e-stamp (constituted of a number of specific properties) by interacting with the IAPP, which provides the necessary stamp information. Subsequently, the email verification process will check the validity of the stamp and submit the email to the mail filtration process. The recipient's end email service updates its dynamic lists according to the outcomes of the mail filtration scheme. This approach has numerous benefits, including the ability to stop illegitimate email from being sent by identifying that there is no adequate IAPP stamp. It also ensures that is verified email is truly legitimate. The challenge here is related to the degree of intrusiveness the mail user is willing to forego in return for added security and legitimacy. Furthermore, the IAPP

functionality introduces additional operational overhead, as well as complexity from a service provisioning perspective.

Klangpraphant and Bhattarakosol employ email authentication to ensure sending legitimacy [2009]. An authentication agent interacts with a user profile that is governed by typical environment security parameters at the operating system level, such as login and password. This is processed by the mail client via a faithful sender agent, which then determines user identity. Broadly speaking, challenge-response techniques rely on the presentation of a *challenge*, such as a password, and expect an associated response, which is computed in realtime or predefined. Various commercial products employ some sort of challenge/response features [Wetzel 2004], and an interesting presending approach based on these principles is presented by Roman et al. [2005]. The sender retrieves the recipient's email and is solicited to interactively supply a reply to a respective challenge before the email is actually sent. This only happens the first time that an email exchange transacts between specific senders and recipients.

The Completely Automatic Public Turing test to tell Computer and Humans Apart (CAPTCHA) [Von Ahn et al. 2004] is similar to general challenge response approaches. CAPTCHA is an elegant technique intended to ensure a degree of confidence that interaction with a specified service is actually being performed by a human being. This degree of confidence depends on the level of logic quality it is based on the work presented in Shirali-Shahreza and Movaghar [2007] and Lin et al. [2004] proposes schemes based on this type of approach. In Shirali-Shahreza and Movaghar [2007], the email user is expected to present specific information to perform the required authentication before being able to send email. The authentication protocol implemented in SASL (Simple Authentication and Security Layer) [SASL 2010a] introduces additional steps in the email-sending process, making it increasingly difficult for spammers to fully automate the act of sending emails in any straight forward fashion. The challenge with this approach is the increased intrusiveness of sending email from an end-user perspective. However, this inconvenience is compensated by the ability to mitigate spam. The approach described in Curran and Honan [2009] employs additional features, including the application of server-side certificates for certifying email servers. A hashcasting technique introduces a stamp intended to identify the validity of the sender's origin and intent by exploiting processing power at the sender's end to generate the stamp. The stamp can be based on a number of sources, for example, SHA-1 [Curran and Honan 2005]. Again, the processing delay introduces a step which renders sending automated mass email much less efficient.

While CAPTCHA-based schemes have become increasingly widespread, challenges remain with respect to the level and degree of intrusiveness such approaches creates for the end user. Similarly, challenge response-based approaches have become increasingly more popular but are faced with a number of comparable challenges, including overall speed degradation of the communication exchange and deadlocking, which occurs when both ends of the communication path are based on challenge responses. To address these challenges, Li et al. [2006] propose an innovative approach: the Mail Transfer Agent (MTA) maintains a trusted list associated with each email user, and when an email's source is not within this trusted list, the sender is challenged. If there is a subsequent reply, the sender is newly added to the recipient's trusted list. To tackle the deadlock issue, the authors propose that when the sender sends an email to a recipient, the recipient's address is automatically added to the sender's trusted list. Additionally, respective trusted lists of both senders and recipients can be modified manually by administrators. This approach involves an extension to the SMTP protocol intended to facilitate the proposed scheme, as well as targets the tempo challenge associated with typical approaches. This introduces complex challenges associated with the protocol itself, which is a widely adopted standard and entrenched in today's email

infrastructure. Any change to it would therefore require considerable effort to ensure the required degree of compatibility. *SpamCooker* is another example that employs a similar approach [Denny et al. 2006]. This technique considerably disturbs spammers' ability to effectively employ traditional techniques for sending unsolicited bulk email. While this is an area which is avidly researched, various studies show that such approaches can still become victims of a number of attacks, which may either, totally bypass or considerably reduce overall effectiveness, depending on the original quality and strength behind the adopted implementations [Bentley and Mallows 2006; Yan and El-Ahmad 2008].

Having a cost element attributed to the process of sending email will harm the spammer's business model, rendering it less favorable. Gansterer et al. [2007] argued that this type of approach is crucial to stopping spam at the source and that privacy-related issues can be less amplified with this method [2006]. LCP (lightweight currency protocol) [Gansterer et al. 2005] and micro-payments [Gansterer et al. 2005; Gansterer et al. 2006; Roman et al. 2005] are typical applications, although concerns regarding potential security issues [Turner and Havey 2004] have been raised. Additionally, there is the complexity of having the necessary logistics to support such an approach, especially when dealing with differences in policies that may exist between service-providing organizations [Gansterer et al. 2005]. The potential variety of platforms may also need to be considered.

4.2. Personalization and Privacy

It would be beneficial to briefly consider the end user's perspective of spam and the challenges it gives rise to. One dimension is that there is an evident element of personal subjectivity: that is, what is considered spam and what is not. Additionally, the level of understanding regarding spam (including how it can be alleviated and what tools and techniques can be employed) varies considerably across the entire email-user base. This is not surprising, given that the number of email users worldwide will reach 1.4 billion in 2009, and 1.9 billion by 2013 [Radicati 2009], amplifiying the importance of putting the end user at the centre of the stage. Therfore, filtering approaches and solutions should seek to be as unobtrusive as possible, to ensure a uniform user experience while allowing for maximum exploitation of spam-reducing opportunities [Lueg and Martin 2007].

As suggested earlier, personalization from an end-user perspective of what is spam and what is not, is an important element that influences the overall perceived value of spam classification [Kiran and Atmosukarto 2009]. Different recipients may use different measures to segregate spam from valid email. It is, therefore, a challenge to ensure that there is a relevant degree of influence from the user to ensure a level of personalization while automating the classification process to the maximum extent possible. This presents a continuing juggling act between the accuracy that can be provided by a small, personal spam-reference baseset, and the more generalized classification provided by a larger baseset, which could potentially lose out on the personalization aspect, assuming both perspectives are considered in isolation. Keeping the personalization aspect as nonintrusive as possible is another challenge, as having the end user manually tag email on a sustained basis (to improve accuracy based on personal considerations) incurs a process overhead that will annoy a number of users.

Junejo and Karim present an approach that specifically considers ensuring personalization in a nonintrusive fashion. [2007]. Based on statistical methods, they employ a more generalized training set for initial setup, which subsequently adapts itself to the end user's context of spam. In the presented perspective, however, the time required for rebuilding statistical models and the sizes of the filters, prove challenging. These process demand significant processing and memory requirements on the Email Service

Provider end and can have a considerable impact where there are a substantial number of relatively large filters and a large number of users. This is one of the reasons why other approaches typically utilize the resources available at the user-level, which is the Mail User Agent end [Hunt and Carpinter 2006].

Tariq et al. also present work on the personal element in spam filtering [2005], employing *Quickfix*, a proposal using a Vector Space Search (VSS), or a word frequency comparison method similar to white-listing approaches. This is employed to compare incoming mail with a localized spam word dictionary. Vector space search is relatively simple and quick, resulting in limited intrusiveness in terms of the entire mail-processing and -sending operation. Subsequent to the VSS operation, if the outcome is marked as spam, the message is forwarded to a Naive Bayesian process. If the outcome of this second operation is still spam, the spam word dictionary is updated accordingly. The authors conclude that while neither VSS nor Bayesian provide an adequate level of consistency in isolation their combination provides an effective approach.

Privacy is another issue that affects the application of spam-filtering approaches, which scrutinizes the email content to establish email legitimacy. This is further amplified when individual third parties other than the sender, intended recipients, and respective service providers are involved. Zhong et al. [2008] discuss the ALPACAS framework, an approach based on a fingerprinting scheme, as a solution. ALPACAS guarantees intactness of the email features while forwarding only a subset of mail content to participating agents for spam classification, rather than a full mail dataset. As a result, the privacy of email can be ensured. There is an assumption that the participating agents are relatively stable; however in reallife scenarios, there are numerous variables which may impact the operation of such an approach and considerably influence the overall value of the framework.

5. DISCUSSIONS

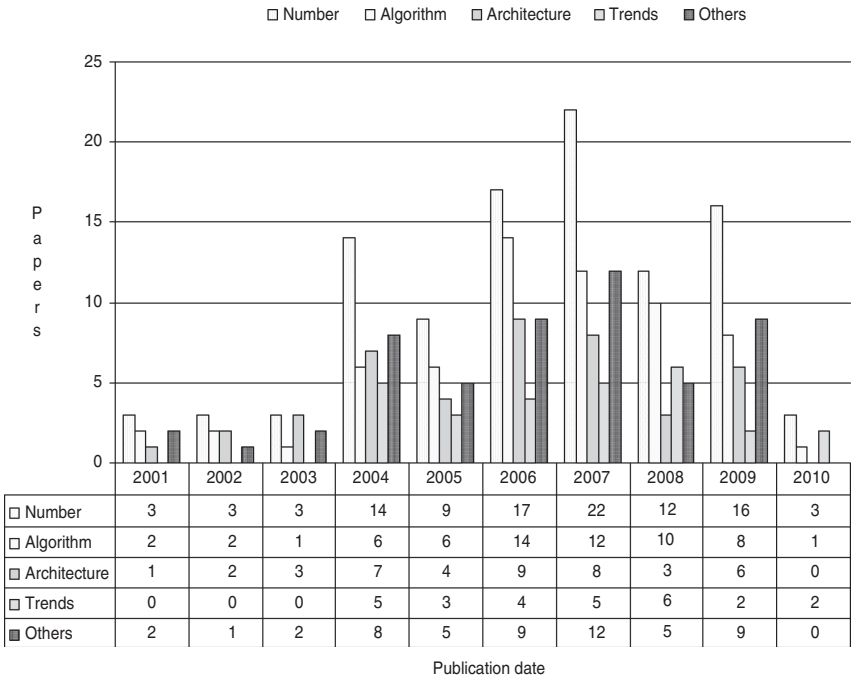
This survey examined a number of emerging approaches to spam filtering, a consulting and reviewing number of sources. Each source varied in scope and covered various aspects of spam filtering techniques.

5.1. Observations

Figure 3 shows the number of papers surveyed, totaling 102 papers ranging from 2001 and 2010 that were surveyed and categorized according to focus, such as Algorithm, Architecture, Trends, and Other. Work classified under Algorithm reflects research that primarily discusses types of classification schemes and associated algorithms, including machine or nonmachine-learning approaches, such as Bayesian, Heuristics, etc. Architecture focused on work principally concerned with the design and implementation of infrastructures—traditional as well as emergent—that enable spam filtering. Work classified under Trends refers to discussions focused on how spam filtering approaches chang over time, which includes the consideration of emerging methods. Other type of research is categorized under the "*Other*" category—here the Works reviewed that could not be directly classified under the other established categories are categorized under other, which includes legal and organizational perspectives, for example.

Figure 4 shows that 35% of surveyed papers focused primarily on the algorithmic perspective. Architectural dimensions were considered at 22%, while about 14% reflected on general trends in anti-spam.

Table I summarizes some perspectives that influence the overall value proposition of a number of spam-filtering technologies, approaches, and techniques. The aspects compared, namely perspective and technique, are referred to as pre- or post-sending filtering approaches. They are also associated with the level of complexity involved,



	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Number	3	3	3	14	9	17	22	12	16	3
Algorithm	2	2	1	6	6	14	12	10	8	1
Architecture	1	2	3	7	4	9	8	3	6	0
Trends	0	0	0	5	3	4	5	6	2	2
Others	2	1	2	8	5	9	12	5	9	0

Fig. 3. A classification of surveyed papers.

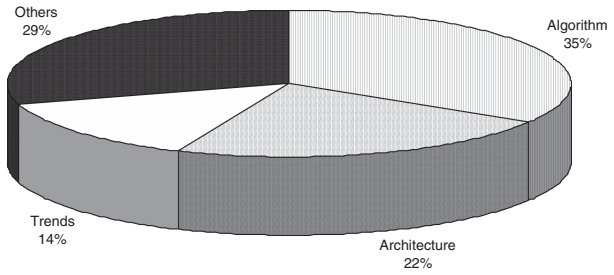


Fig. 4. The focus distribution of the surveyed papers.

overall performance, and quality, as well as how proliferated popular they are, in terms of *Reach*.

Table I assigns an identifier to each dimension demonstrating whether the approach employs machine-learning (ML) techniques or otherwise (Non-ML). The table also shows whether distributed and collaborative elements are in use. The last two columns in Table I illustrate the levels of intrusiveness and privacy-preserving from an end-user perspective.

Rule, signature, and Bayesian approaches have been widely discussed, and the overall, long-term value of the latter approach tends to be better than the first two. However, while signature- and rule-based approaches are generally considered less intrusive, Bayesian-based classification remains highly popular [Khorsi 2007; Song et al. 2009; Chen et al. 2008; Yang et al. 2006]. While performance and requirements vary considerably, scalability is an ongoing concern for a number of traditional Bayesian-based approaches, due to their dependence on memory availability and performance [Li and

Table I. Surveyed Work Perspectives

Perspective	Technique	Pre-Send	Post-Send	Complexity	Performance	Basic Quality	Reach	ML	Non-ML	Distributed	Collaborative	Privacy	Intrusiveness
Algorithm	Rule Based	√	√	⇓	⇓	⇔	⇑	x	√	O	O	⇓	⇓
Algorithm	Signature	√	√	⇓	⇔	⇔	⇑	x	√	O	O	⇓	⇓
Algorithm	Bayesian	x	√	⇔	⇔	⇑	⇑	√	x	O	O	⇔	⇔
Algorithm	kNN	x	√	⇔	⇔	⇔	⇔	√	x	O	O	⇔	⇔
Algorithm	ANN	x	√	⇔	⇔	⇔	⇔	√	x	O	O	⇔	⇔
Algorithm	SVM	x	√	⇑	⇑	⇑	⇓	√	x	O	O	⇔	⇔
Method	Monetary	√	x	⇔	⇑	⇑	⇓	O	O	O	O	⇑	⇑
Method	Challenge/Response	√	x	⇔	⇑	⇑	⇓	O	O	O	O	⇑	⇑
Method	Semantic	x	√	⇔	⇔	⇔	⇓	O	O	√	√	⇓	⇔
Method	Social Net.	x	√	⇔	⇔	⇔	⇓	O	O	√	√	⇓	⇔
Method	Infrastructure	√	√	⇑	⇔	⇔	⇓	O	O	O	O	⇑	⇓
Architecture	GRID	x	√	⇔	⇔	⇑	⇓	O	O	√	√	⇓	⇔
Architecture	P2P	x	√	⇔	⇔	⇑	⇔	O	O	√	√	⇓	⇔

√	x	⇓	⇑	⇔	O
yes	no	low	high	average	not in text

Zhong 2006]. Spammers can also learn which terms influence Bayesian-filtering outcomes and adopt an approach based on the additional insertion of spurious content that is intended to influence overall results and outcomes. This is accomplished by weakening the effect of any high-order ranked tokens used to infer whether the content is spam.

Other common machine-learning approaches include Artificial Neural Networks (ANN), k Nearest Neighbor (kNN), and, more recently, Support Vector Machines (SVMs). kNN approaches are normally subjective to noise, indicating that errors in the training set can easily induce misclassification. They also tend to be computationally intensive in terms of larger datasets. While the accuracy of ANN classification is high, this approach has a tendency to require significant computing time for spam classification [Gansterer et al. 2006]. In online anti-spam classification environments, the balance between having an up-to-date training dataset and resources available to train or retrain the artificial neural network is critical. This requires a continued effort of striking a balance between functionality and effectiveness. It is common to identify real-world scenarios ANN-based approaches are used in conjunction with additional filtering schemes in which rather than in isolation. SVM approaches have shown their effectiveness in spam filtering. Classification in a SVM approach is founded on the notion of *hyperplanes*, which acts as class segregators [Scholkopf and Smola 2001]. The SVM's goal is primarily to identify the best possible hyperplane in the context of the selection of the largest distance (or margin) between the closest representative points referred to as *support vectors*. In contrast to Bayesian approaches, SVM approaches are even more computationally expensive, which to constrains their maximum potential application in online implementations.

Filtering approaches, although not as popular as machine-learning approaches, are based on challenge response, authentication, and CAPTCHA schemes, and tend to provide a higher degree of performance in terms of the ability to mitigate spam proliferation. They also ensure a relatively higher degree of privacy and security but can be more intrusive in terms of overall user experience. From a holistic architectural perspective, numerous methods in application have been identified. For the purposes of simplicity, these can be grouped under a number of umbrella areas, namely Mail Transfer Agent

Table II. Typical Products and Types

Product	Spam Filtering Type
Google/Yahoo/Hotmail	Hosted
McAffee/SpamKiller	Appliance
NetIQ/BrightMail/GFI Essentials	Centralized Gateway
NetworkBox	IPS/IDS
IronPort/Barracuda/SpamTitan	Appliance
BitDefender	MTA Extension
SpamAssassin	Software
Cloudmark SafetyBar/Spam Bayes/Outclass	MUA Plugin

(MTA), Mail User Agent (MUA), Hosted, P2P-computing, and grid computing-based approaches. In MTA, the actual implementation is commonly identified as either an extension to the MTA or part and parcel of the package. The MUA approach commonly involves rules which can be programmed within the MUA environment itself. They also include plugins with filtering intelligence, as well as external solutions, which work outside the main MUA. SMTP proxy services, usually available as separate stand alone logic, are also popular. Under normal circumstances, the hosted approach does not involve any (or perhaps minimal) intelligence from the MUA or MTA perspective, as the entire (or most of the) anti-spam filtering service is performed elsewhere.

P2P, and grid computing-based paradigms in which participating nodes are able to share various resources such as storage, processing power, and connectivity in a collaborative way, are considered as emerging, high-performance spam filtering schemes. High resiliency is a key advantage of such architectures, but their subjectivity to participating nodes with malicious intent is of concern. Another benefit of these approaches is their ability to share spam intelligence relatively easily, widening the scope for increased collaboration. The same applies to social network and semantic Web-based approaches, which are also garnering increased attention. These emerging approaches, however, are not currently widespread compared with traditional approaches.

Other approaches exist as well. It is worth mentioning that specialized devices that are positioned strategically within an enterprise network to minimize spam influx are increasingly common. Such devices include intrusion and prevention detection systems (IPD/S), network devices which are either enabled out-of-the-box with specific spam filtering functionality or which can be extended to enable such functionality, as well as specialized devices specifically and solely intended to act as spam filtering devices. Table II provides a market snapshot of popular solutions to spam filtering.

Overall, it has been identified is that no one size fits all. In the context of the methods, algorithms, and architectures employed to mitigate the spam challenge, the selection and combination of the techniques discussed have varying outcomes that influence their overall success in terms of adoption and implementation. In isolation, performance figures such as speed and accuracy are not necessarily inductive to the most effective approaches in the real-world. Architecture brilliance doesn't necessarily create the most effective environment overall for spam filtering. The consideration of other important factors comes into play, including the complexity of implemention, end-user intrusiveness, and privacy amongst others.

The inability to identify a singular superlative approach towards spam filtering is one of the driving focus that continues to stimulate sustained research catalysis for identifying better ways for mitigating its proliferation.

5.2. Research Directions

Spam can be considered an Internet-scale problem, therefore demanding high-performance scalable algorithms. Through this survey work, a potential spam-fittering research opportunity has been identified: the decoupling of key resources, that is, CPU,

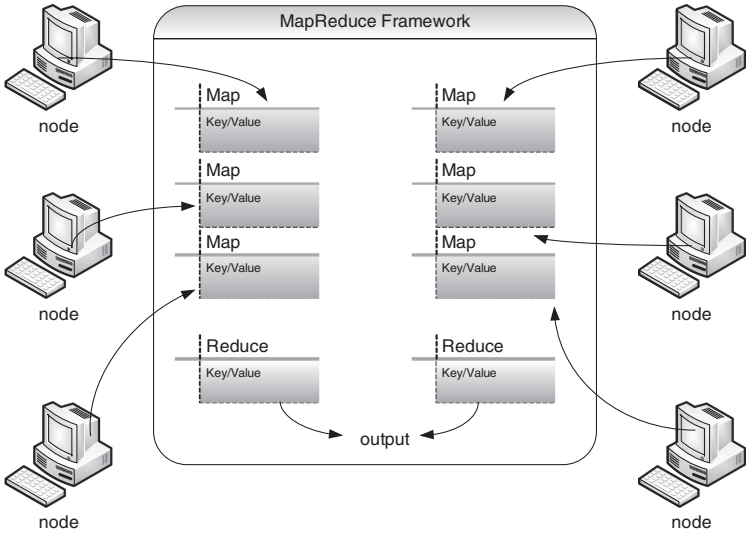


Fig. 5. The MapReduce framework.

memory, and storage, in distributed systems, facilitates parallelization of algorithms. Distributed systems exhibit a number of intrinsic properties making them very suitable candidates for tackling large-scale computing problems. In this respect, distributed, and parallel-computing principles are believed to be key enablers for providing intelligent Internet-scale anti-spam solutions.

While parallel programming in distributed environments is not a trivial task, various frameworks and models are available to make distributed and parallel computing more accessible. The MapReduce framework is a programming model intended to abstract large-scale computation challenges and enable automatic parallelization [Dean and Ghemawat 2008]. MapReduce was popularized by Google and primarily motivated by the need to parallelize the processing of Internet-scale datasets. Programmatically inspired from functional programming, there are two primary features at its core namely a *map* and a *reduce* operation. From a logical perspective, all data is treated as a Key (K) and Value (V) pair. Multiple mappers and reducers are employed at an atomic level; however, a map operation takes a $\{K_1, V_1\}$ pair and emits an intermediate list $\{K_2, V_2\}$ pairs. A reduce operation takes all values represented by the same key in the intermediate list and processes them accordingly, emitting a final new list $\{V_2\}$.

Figure 5 provides a high-level pictorial representation of a typical MapReduce implementation.

The consideration for and application of MapReduce-based approaches in the context of distributed systems is hardly new, but it is also an area which one can safely assert is relatively fresh in terms of exploiting its potential for application in specialized ways. This includes uniting it with modern machine-learning algorithms for spam detection and filtering, such as support vector machines [Sculley and Wachman 2007]. To date, the primary application of MapReduce is toward data-intensive tasks, rather than for computation in general, efforts do exist, including the work presented in Das et al. [2007], Dean [2006], Dean and Ghemawat [2008], and Noll and Meinel [2008].

MapReduce can be employed for dealing with both data and computationally intensive challenges. It also provides a motivating research prospect, due to the specific characteristics associated with this (spam filtering) type of problem, involving searching and analyzing specific patterns within collectively large corpuses of data. Tuning

contemporary and innovative machine-learning-based spam-filtering techniques to this metaphor will provide ample potential for research.

In the application of a distributed anti-spam filtering architecture, and in the context of MapReduce, respective *Mappers* and *Reducers* will perform the steps required for machine-learning computing sequences. Distributing spam-filtering jobs and spam data to a potentially unlimited number of computing resources can effectively tackle the spam challenge in a large-scale context. MapReduce can be employed as an enabling technology to facilitate high-performance machine-learning techniques for spam filtering. For example, SVM training is well known to be a compute-intensive task because of the quadratic programming challenge it involves, that is, training time increases exponentially with the number of training elements involved. Distributed MapReduce-based SVM training would scale well as a number of nodes could be utilized and each node could deal with a subset of the training data at an individual level, rather than processing the entire dataset, thus mitigating this problem.

6. CONCLUSIONS AND FUTURE WORK

Spam is an Internet-scale problem. Its proliferation has reached notable proportions. Approaches to spam filtering have been continuously researched and explored with varying degrees of success. The techniques applied and employed by spammers continue to get smarter, with the primary intent being to outsmart counterpart detection and filtering schemes.

This survey focused on emerging approaches to spam filtering. Based on the research performed and literature reviewed, it appears that capitalization of resources at the client's end improves the scope for increasing anti-spam intelligence [Zhong et al. 2008; Liu et al. 2008; Noll and Meinel 2008; Dimmock and Maddison 2004; Kong et al. 2006; Brodsky and Brodsky 2007]. Combinations of filtering schemes also provide better overall results when compared with singular approaches [Lynam et al. 2006]. Emerging approaches based on grid computing, P2P computing, semantic Web, and social networks bring increased opportunities in terms of scalability, formalization of spam definitions, and personalization, and collaboration between the participating parties by sharing resources and increasing spam intelligence.

It is noted that a good number of modern-day email-service providers still rely primarily on centralized services. These approaches mostly use traditional classification schemes for spam-filtering efforts, while striving to ensure an adequate level of personalization. This popularity can be attributed to the simplicity and non-intrusiveness of the approach. On the other hand, the costs associated with serving a large population of end users using a rigid centralized function cannot be ignored. The elements associated with costs concern not only the financial aspects, but encompass the processing aspect, including classification scheme performance and requirements, and storage and network perspectives. Fluctuation of user population, both in terms of concurrency, as well as their number, also has a considerable impact.

Spammers are getting smarter, continuously trying to come up with approaches that enable them to circumnavigate spam-filtering schemes. Furthermore, active research in emerging approaches shows that there is a continued effort to come up with better and alternative anti-spam schemes. Given that the operational landscape of spam ecosystems is continuously changing, different techniques that can be applied to old and new problem areas are being sought. Social networks, for example, have become a major breeding ground for spam-related activities. Research in these specific areas, including the behavioral model space [Wu 2009; Ramachandran and Feamster 2006], has also intensified with continued advancement.

Considering the large scale of the spam challenge, this survey also pointed out that distributed computing paradigms, such as the MapReduce-framework, can be employed

as an enabling technology for high-performance machine-learning approaches to spam filtering. This rationale is considered as a baseline proposition for future work to research the feasibility of a flexible, collaborative, parallel, and distributed spam-filtering architecture. Such architecture is intended to scale at the rate mandated by any potential increase and decrease of spam. Rather than focusing on classification schemes or underlying architectures in isolation, an in-depth level of consideration is applied to both dimensions concurrently, with the intention of providing a synergistic scheme able to exploit the best of both dimensions at the same time. This collaborative notion can extend to sharing of spam intelligence achieved by involving end users, service providers, private and public sector institutions. A longterm objective can be the suggestion and realization of an institutionalized global virtual organization supporting a global spam-filtering ecosystem.

Finally, the quality, as well as the number of related literature, demonstrates the effort and significant advancements that have been, and continue to be made regarding the spam challenge.

REFERENCES

ALLMAN, E. 2006. Email authentication: what, why, how? *ACM Queue, 4,* 9, 30–34.

ANTONOPOULOS, A., ALEXANDROS, G., STEFANIDIS, K., AND ARTEMIOS G. V. 2009. Fighting spammers with spam. In *Proceedings of the 9th IEEE International Symposium on Autonomous Decentralized Systems (ISADS'09)*. http://www.artemiosv.info/Spam.pdf.

BARRENO, M., NELSON, B., SEARS, R., JOSEPH, A. D., AND TYGAR, J. D. 2006. Can machine learning be secure? In *Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASI-ACCS'06)*. ACM, New York, NY, 16–25.

BENEVENUTO, F., RODRIGUES, T., ALMEIDA, V., ALMEIDA, J., ZHANG, C., AND ROSS, K. 2008. Identifying video spammers in online social networks. In *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb'08).* ACM Press, New York, NY, 45–52.

BENTLEY, J. AND MALLOWS, C. 2006. CAPTCHA challenge strings: problems and improvements. In *Proceedings of the Document Recognition & Retrieval*. 141–147.

BERNSTEIN, D. 2000. Internet Mail 2000. http://cr.yp.to/im2000.html.

BLANZIERI, E. AND BRYL, A. 2008. A survey of learning-based techniques of email spam filtering. *Artif. Intell. Rev.* 29, 1, 63–92.

BOYD, D. M. AND ELLISON, N. B. 2007. Social network sites: definition, history, and scholarship. *J. Comput. Mediated Comm. 13*, 1, 210–230.

BRIN, S. AND PAGE, L. 1998. The anatomy of a large-scale hypertextual Web search engine. *Comput. Netw*. 30, 1-7, 107–117.

BRODSKY, A. AND BRODSKY, D. 2007. Trinity: distributed defense against transient spam-bots. In *Proceedings of the 26th Annual ACM Symposium on Principles of Distributed Computing*. ACM, New York, NY, 378–379.

BROWN, G., HOWE, T., IHBE, M., PRAKASH, A., AND BORDERS, K. 2008. Social networks and context-aware spam. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work.* ACM, New York, NY, 403–412.

CALAIS, P., GUEDES, D., MEIRA JR., W., HOEPERS, C., CHAVES, M., AND JESSEN, K. S. 2009. Spamming chains: a new way of understanding spammer behavior. In *Proceedings of the 6th Conference on Email and Anti-Spam (CEAS)*.

CHEN, C., TIAN, Y., AND ZHANG, C. 2008. Spam filtering with several novel bayesian classifiers. In *Proceedings the 19th IEEE International Conference on Pattern Recognition.* 1–4.

CHEN, J., XIAO, G., GAO, F., AND ZHANG, Y. 2008. Spam filtering method based on an artificial immune system. In *Proceedings of the IEEE International Conference on Multimedia and Information Technology (MMIT)*. 169–171.

CHIRITA, P., DIEDERICH, J., AND NEJDL, W. 2005. MailRank: using ranking for spam detection. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management (CIKM'05)*. ACM, New York, NY, 373–380.

CHORD. 2009. Peer-to-Peer Lookup. http://pdos.csail.mit.edu/chord/faq.html.

CONRY-MURRAY, A. 2004. The antispam cocktail: Mix it up to stop junk email. *Netw. Mag. 19*, 7, 33–38.

COOK, D., HARTNETT, J., MANDERSON, K., AND SCANLAN, J. 2006. Catching spam before it arrives: Domain specific dynamic blacklists. In *Proceedings of the Australasian Workshops on Grid Computing and e-Research*, vol. 54, Australian Computer Society, Inc. Darlinghurst, Australia, 193–202.

CORMACK, G. V. AND LYNAM, T. R. 2007. Online supervised spam filter evaluation. *ACM Trans. Inf. Syst. 25*, 3, Article 11.

CURRAN, K. AND HONAN, J. S. 2005. Addressing spam e-mail using hashcast. *Inter. J. Bus. Admin. 1*, 2, 41–65.

DAMIANI, E., VIMERCATI, S., PARABOSCHI, S., AND SAMARATI, P. 2004. P2P-based collaborative spam detection and filtering. In *Proceedings of the 4th IEEE International Conference on Peer-to-Peer Computing (P2P'04)*.

DAS, A., DATAR M., GARG, A., AND RAJARAM, S. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*. ACM, New York, NY, 271–280.

DEAN, J. AND GHEMAWAT, S. 2008. MapReduce: Simplified data processing on large clusters. *Commun. ACM 51*, 1, 107–113.

DEAN, J. 2006. Experiences with MapReduce, an abstraction for large-scale computation. In *Proceedings of the 15th International Conference on Parallel Architectures and Compilation Techniques (PACT'06)*. ACM, New York, NY, 1–1.

DENNY, N. El HOURANI, T., DENNY, J., BISSMEYER, S., AND IRBY, D. 2006. SpamCooker: A method for deterring unsolicited electronic communications. In *Proceedings of the 3rd International Conference on Information Technology: New Generations (ITNG'06)*. Los Alamitos, CA, 590–591.

DIMMOCK, N. AND MADDISON, I. 2004. Peer-to-peer collaborative spam detection. *Crossroads Mag. 11*, 2, 4–4.

DNSBL, 2009. The Spam Database Lookup. `http://www.dnsbl.info/`.

DUAN, Z., DONG, Y., AND GOPALAN, K. 2007. DMTP: controlling spam through message delivery differentiation. *Comput. Netw. 51*, 10, 2616–2630.

ESQUIVEL, H., MORI, T., AND AKELLA, A. 2009. RouterLevel spam filtering using TCP fingerprints: architecture and measurement based evaluation. In *Proceedings of the 6th Conference on Email and Anti-Spam*.

FERRIS RESEARCH. 2008. Industry Statistics-Ferris Research. `http://www.ferris.com/research-library/industry-statistics/`.

FOUKIA, N. ZHOU, L., AND NEUMAN, B. C. 2006. Multilateral decisions for collaborative defense against unsolicited bulk email. In *Proceedings of the 4th International Conference on Trust Management*, *Lecture Notes in Computer Science*. 77–92.

GANSTERER, W. N. AND ILGER, M. 2007. Analyzing UCE/UBE traffic. In *Proceeding of the 9th International Conference on Electronic Commerce (ICEC'07)*. ACM, New York, NY, 195–204.

GANSTERER, W., HLAVACS, H., ILGER, M., LECHNER, P., AND STRAUB, J. 2006. Token buckets for outgoing spam prevention. In *Proceedings of the IASTED International Conference on Communication, Network, and Information Security*. 36–41.

GANSTERER, W., ILGER, M., LECHNER, P., NEUMAYER, R., AND STRAUB, J. 2005. Anti-spam methods: state of the art. Techn. rep., FA384018-1, Institute of Distributed and Multimedia Systems, University of Vienna.

GARG, A., BATTITI, R., AND CASCELLA, R. G. 2006. May I borrow your filter? Exchanging filters to combat spam in a community. In *Proceedings of the 20th International Conference on Advanced Information Networking and Applications*. Los Alamitos, CA, 489–493.

GBURZYNSKI, P. AND MAITAN, J. 2004. Fighting the spam wars: a remailer approach with restrictive aliasing. *ACM Trans. Internet Technol. 4,* 1, 1–30.

GFI, 2010. Hosted spam filtering service. `http://www.gfi.com/landing/maxmp-hosted-spam-filtering-service.asp?adv=69&loc=643&gclid=CJCy9bqVuKACFUmS3wodhFQaTQ`.

GNUTELLA. 2001. The gnutella protocol specification. `http://www9.limewire.com/developer/gnutella_protocol_0.4.pdf`.

GOMES, L. H., CAZITA, C., ALMEIDA, J. M., ALMEIDA, V., AND MEIRA, J. 2004. Characterizing a spam traffic. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement (IMC'04)*. ACM, New York, NY, 356–369.

GOODMAN, J., CORMACK, G. V., AND HECKERMAN, D. 2007. Spam and the ongoing battle for the inbox. *Commun. ACM 50*, 2, 24–33.

GRAY, A. AND HAAHR, M. 2004. Personalized, collaborative spam filtering. In *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS)*.

GUZELLA, T. S. AND CAMINHAS, W. M. 2009. A review of machine learning approaches to spam filtering. *Expert Syst. Appl. 36*, 7, 10206–10222.

HAYATI, P. AND POTDAR, V. 2008. Evaluation of spam detection and prevention frameworks for email and image spam: a state of art. In *Proceedings of the 10th International Conference on Information Integration and Web-Based Applications & Services (iiWAS'08)*. ACM, New York, NY, 520–527.

HERSHKOP, S. AND STOLFO, S. J. 2005. Combining email models for false positive reduction. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD'05)*. ACM, New York, NY, 98–107.

HIDALGO. 2005. Machine learning for spam detection references. `http://www.esi.uem.es/~jmgomez/spam/MLSpamBibliography.html`.

HUNT, R. AND CARPINTER, J. 2006. Tightening the net: a review of current and next generation spam filtering tools. *Comput. Security, 25*, 566–578.

HSIA, J. AND CHEN, M. 2009. Language-model-based detection cascade for efficient classification of image-based spam e-mail. In *Proceedings of the IEEE international Conference on Multimedia and Expo (ICME'09)*. IEEE Press, Los Alamitos, CA, 1182–1185.

JANECEK, A. G., GANSTERER, W. N., AND KUMAR, K. A. 2008. Multi-level reputation-based greylisting. In *Proceedings of the 3rd International Conference on Availability, Reliability and Security (ARES'08)*. Los Alamitos, CA, 10–17.

JUNEJO, K. N. AND KARIM, A. 2007. PSSF: A novel statistical approach for personalized service-side spam filtering. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)*. Los Alamitos, CA, 228–234.

JUNG, J. AND SIT, E. 2004. An empirical study of spam traffic and the use of DNS black lists. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement (IMC'04)*. ACM, New York, NY, 370–375.

KHANAL, A., MOTLAGH, B. S., AND KOCAK, T. 2007. Improving the efficiency of spam filtering through cache architecture. In *Proceedings of the 15th international Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS'07)*. Los Alamitos, CA, 303–309.

KHORSI, A. 2007. An overview of content-based spam filtering techniques. *Informatica 31*, 269–277.

KIM, J., DOU, D., LIU, H., AND KWAK, D. 2007. Constructing a user preference ontology for anti-spam mail systems. In *Proceedings of the 20th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence (CAI'07)*. 272–283.

KIRAN, P. AND ATMOSUKARTO, I. 2009. Spam or Not Spam—That is the question. Tech. rep., University of Washington. `http://www.cs.washington.edu/homes/indria/research/spamfilter_ravi_indri.pdf`.

KOLCZ, A. AND YIH, W. 2007. Raising the baseline for high-precision text classifiers. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*. ACM, New York, NY, 400–409.

KONG, J. S., REZAEI, B. A., SARSHAR, N., ROYCHOWDHURY, V. P., AND BOYKIN, P. O. 2006. Collaborative spam filtering using e-mail networks. *Computer, 39*, 8, 67–73.

LEE, T. B., HENDLER, J., AND LASILLA, O. 2001. The semantic web. *Scientific Amer. 284*, 5, 34–43.

LEVINE, J. R. 2005. Experiences with greylisting. In *Proceedings of the 2nd Conference on Email and Anti-Spam (CEAS)*.

LAWTON, G. 2005. Email authentication is here, but has it arrived yet? *Computer, 38*, 11, 17–19.

LI, M. AND BAKER, M. 2005. *The Grid Core Technologies*. John Wiley & Sons.

LI, Q. AND MU, B. 2009. A novel method to detect junk mail traffic. In *Proceedings of the 9th International Conference on Hybrid Intelligent Systems*. Los Alamitos, CA, 129–133.

LI, K. AND ZHONG, Z. 2006. Fast statistical spam filter by approximate classifications. In *Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS'06/Performance'06)*. ACM, New York, NY, 347–358.

LIANG, J., KUMAR, R., XI, Y., AND ROSS, K. W. 2005. Pollution in P2P file sharing systems. In *Proceedings of IEEE International Conference on Computer Communications (INFOCOM'05)*. 1174–1185.

LIAO, L. AND SCHWENK, J. 2007. End-to-end header protection in signed S/MIME. In *Proceedings of the OTM Confederated International Conference on On the Move to Meaningful Internet Systems*. 1646–1658.

LIMEWIRE. 2009. Peer To Peer & Torrent Client. `http://www.limewire.com`.

LIN, P. L., HUANG, J. L., AND CHANG, T. J. 2004. CAPTCHA-based anti-spam mail model. In *Proceedings of the International Conference on Security and Management (SAM'04)*. 332–338.

LIU, P. SHI, Y. F. LAU, F. AND CHO-LI, W. 2005. Anti-spam grid: a dynamically organized spam filtering infrastructure. In *Proceedings of the 5th WSEAS International Conference on Simulation, Modeling and Optimization*. 61–66.

LUO, P., XIONG, H., LÜ, K., AND SHI, Z. 2007. Distributed classification in peer-to-peer networks. In *Proceedings of the 13th ACM SIGKDD International conference on Knowledge Discovery and Data Mining (KDD'07)*. ACM, New York, NY, 968–976.

LUEG, C. AND MARTIN, S. 2007. Users dealing with spam and spam filters: Some observations and recommendations. In *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction: Design Centered HCI (CHINZ'07)*. ACM, New York, NY, 67–72.

LYNAM, T. R., CORMACK, G. V., AND CHERITON, D. R. 2006. On-line spam filter fusion. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06)*. ACM, New York, NY, 123–130.

MA, W., TRAN, D., AND SHARMA, D. 2009. A novel spam email detection system based on negative selection. In *Proceedings of the 4th International Conference on Computer Sciences and Convergence Information Technology (ICCIT'09)*. Los Alamitos, CA, 987–992.

METZGER, J., SCHILLO, M., AND FISCHER, K. 2003. A multiagent-based peer-to-peer network in java for distributed spam filtering. In *Proceedings of the 3rd Central and Eastern European Conference on Multi-Agent Systems (CEEMAS'03)*. 616–625.

MICROSOFT. 2010. Microsoft online services-hosted exchange services. `http://www.microsoft.com/online/exchange-hosted-services.mspx`.

NOLL, M. G. AND MEINEL, C. 2008. Building a scalable collaborative web filter with free and open source software. In *Proceedings of the IEEE International Conference on Signal Image Technology and Internet Based Systems (SITIS'08)*. Los Alamitos, CA, 563–571.

OHFUKU, H. AND MATSUURA, K. 2006. Integration of Bayesian filter and social network technique for combating spam e-mails better. *Trans. Info. Process. Soc. Japan 47*, 8, 2548–2555.

ORGUN, B., DRAS, M., NAYAK, A., AND JAMES, G. 2006. Approaches for semantic interoperability between domain ontologies. In *Proceedings of the Australasian Ontology Workshop (AOW'06)*. 41–50.

PATHAK, A., HU, Y. C., AND MAO, Z. M. 2008. Peeking into spammer behavior from a unique vantage point. In *Proceedings of the 1st USENIX Workshop on Large-Scale Exploits and Emergent (LEET'08)*. Article 3.

PAUL, P. P., JUDGE, P., ALPEROVITCH, D., AND YANG, W. 2005. Understanding and reversing the profit model of spam. In *Proceedings of the Workshop on the Economics of Information Security (WEIS)*.

PETERSON, P. 2006. E-mail is broke! Authentication can fix it. `http://electronicdesign.com/article/embedded/e-mail-is-broke-authentication-can-fix-it12463.aspx`.

RADICATI, 2009. Email statistics report, 2009–2013. `http://www.radicati.com/wp/wp-content/uploads/2009/05/email-stats-report-exec-summary.pdf`.

RAMACHANDRAN, A. AND FEAMSTER, N. 2006. Understanding the network-level behavior of spammers. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'06)*. ACM, New York, NY, 291–302.

RAMACHANDRAN, A., FEAMSTER, N., AND VEMPALA, S. 2007. Filtering spam with behavioral blacklisting. In *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS'07)*. ACM, New York, NY, 342–351.

ROWSTRON, A. AND DRUSCHEL, P. 2001. Pastry: scalable, distributed object location and routing for large-scale peer-to-peer systems. In *Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms (Middleware'01)*, Heidelberg, Germany, Lecture Notes in Computer Science, vol. 2218, Springer, 329–350.

ROMAN, R., ZHOU, J., AND LOPEZ, J. 2005. Protection against spam using pre-challenges. In *Proceedings of the 20th International Conference on Information Security*. 281–294.

SALTON, G. AND BUCKLEY, C. 1988. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage. 24*, 5, 513–523.

SASL. 2010. IETF. RFC 2222. `http://www.ietf.org/rfc/rfc2222.txt`.

SCHOLKOPF, B. AND SMOLA, A. J. 2001. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press, Cambridge, MA.

SCHOLLMEIER, R. 2001. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In *Proceedings of the 1st IEEE International Conference on Peer-To-Peer*. 101–102.

SEBASTIANI, F. 2002. Machine learning in automated text categorization. *ACM Comput. Surv. 34*, 1, 1–47.

SCULLEY, D. AND WACHMAN, G. M. 2007. Relaxed online SVMs for spam filtering. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07)*. ACM, New York, NY, 415–422.

SECUREWORKS. 2009. Spam botnets to watch in 2009. `http://www.secureworks.com/research/threats/botnets2009/`.

SHIRALI-SHAHREZA, S., AND MOVAGHAR, A. 2007. A new anti-spam protocol using CAPTCHA. In *Proceedings of the International Conference on Computer Systems and Applications*. 234–238.

SONG, Y., KOLCZ, A., AND GILES, C. L. 2009. Better naive bayes classification for high-precision spam detection. *Softw. Pract. Exper. 39,* 11, 1003–1024.

SPAMASSASIN. 2009. The Apache SpamAssassin project. `http://spamassassin.apache.org/`.

SpamCop. 2009. http://www.spamcop.net/.

Spamhaus. 2009. The Spamhaus project. http://www.spamhaus.org/.

Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., and Balakrishnan, H. 2001. Chord: A scalable peer-to-peer lookup service for internet applications. *SIGCOMM Comput. Commun. Rev. 31*, 4, 149–160.

The Register. 2004. Spammers embrace email authentication. http://www.theregister.co.uk/2004/09/03/email_authentication_spam/.

Turner, D. and Havey, D. 2004. Controlling spam through lightweight currency. In *Proceedings of the International Conference on Computer Sciences*.

Twining R. D., Williamson M. M., Mowbray M., and Rahmouni M. 2004. Email prioritization: reducing delays on legitimate mail caused by junk mail. Tech. rep., HPL-2004-5R1, HP Labs. http://www.hpl.hp.com/techreports/2004/HPL-2004-5R1.pdf.

von Ahn, L., Blum, M., and Langford, J. 2004. Telling humans and computers apart automatically. *Comm. ACM 47*, 2, 56–60.

Wei, C., Sprague, A., and Warner, G. 2009. Clustering malware-generated spam emails with a novel fuzzy string matching algorithm. In *Proceedings of the ACM Symposium on Applied Computing (SAC'09)*. ACM, New York, NY, 889–890.

Weiss, G. M. and Tian, Y. 2006. Maximizing classifier utility when training data is costly. *SIGKDD Explor. Newsl. 8*, 2, 31–38.

Wetzel, R. 2004. Spam fighting business models-who wins, who loses. *Bus. Comm. Rev. 34*, 24–9.

Witten, H. and Eibe, F. 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd Ed., Morgan Kaufmann, San Francisco.

Wu, C. 2009. Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks. *Expert Syst. Appl. 36*, 3, 4321–4330.

Xie, Y., Yu, F., Achan, K., Panigrahy, R., Hulten, G., and Osipkov, I. 2008. Spamming botnets: Signatures and characteristics. In *Proceedings of the ACM Conference on Data Communication (SIGCOMM)*. ACM, New York, NY, 171–182.

Yan, J. and El-Ahmad, A. S. 2008. A low-cost attack on a microsoft CAPTCHA. In *Proceedings of the 15th ACM Conference On Computer And Communications Security (CCS'08)*. ACM, New York, NY, 543–554.

Yang, Z. Nie, X. Xu, W., and Guo, J. 2006. An approach to spam detection by naive Bayes ensemble based on decision induction. In *Proceedings of the 6th IEEE International Conference on Intelligent Systems Design and Applications*. 861–866.

Yang, Y. and Pedersen, J. O. 1997. A comparative study on feature selection in text categorization. In *Proceedings of the 14th International Conference on Machine Learning (ICML'97)*. 412–420.

Yeh, C., Mao, C., Lee, H., and Chen, T. 2007. Adaptive e-mail intention finding mechanism based on e-mail words social networks. In *Proceedings of the Workshop on Large Scale Attack Defense (LSAD'07)*. ACM, New York, NY, 113–120.

Youn, S. and McLeod, D. 2009a. Spam decisions on gray e-mail using personalized ontologies. In *Proceedings of the ACM Symposium on Applied Computing (SAC'09)*. ACM, New York, NY, 1262–1266.

Youn, S. and Mcleod, D. 2009b. Improved spam filtering by extraction of information from text embedded image e-mail. In *Proceedings of the ACM Symposium on Applied Computing (SAC'09)*. ACM, New York, NY, 1754–1755.

Yuan, Z., Zhang, Q., Li, D., and Liu, Y. 2007. Research of anti-spam application architecture based on semantic grid. In *Proceedings of the 2nd International Conference on Innovative Computing, Information and Control*. Los Alamitos, CA, 491–491.

Zhang, L., Zhu, J., and Yao, T. 2004. An evaluation of statistical spam filtering techniques. *ACM Trans. Asian Lang. Inf. Process. 3*, 4, 243–269.

Zhong, Z., Ramaswamy, L., and Li, K. 2008. ALPACAS: A large-scale privacy-aware collaborative anti-spam system. In *Proceedings of the 27th Conference on Computer Communications*. 556–564.

Zhou, F., Zhuang, L., Zhao, B. Y., Huang, L., Joseph, A. D., and Kubiatowicz, J. 2003. Approximate object location and spam filtering on peer-to-peer systems. In *Proceedings of the ACM/IFIP/USENIX International Conference on Middleware (Middleware'03)*. Verlag, Berlin, 1–20.

Zinman, A. and Donath, J. 2007. Is Britney Spears spam? In *Proceedings of the 4th International Conference on Email and Anti-Spam (CEAS)*.

Received November 2009; revised March 2010; accepted July 2010