

Classifying Excessive Internet Resource Consumption – Proposed Taxonomies

Minh Tran, Grenville Armitage
Centre for Advanced Internet Architectures
Swinburne University of Technology
Melbourne, Australia
{mtran,garmitage}@swin.edu.au

Abstract- The Internet’s current architecture de-couples the cost of sending traffic from the consequences of sending traffic. Its open nature allows anyone to initiate a connection with minimal authentication. These factors account for the excessive consumption caused by spam in the email system and other service areas. At the network layer, resource consumption caused by spam is analogous to Distributed Denial-of-Service (DDoS) attacks. We believe the two taxonomies proposed in this paper (of spam definition and of spam sending mechanisms) are the first comprehensive study of spam over different service areas. Our taxonomies will help in comprehending the similarities and differences between traditional Email spam and related newer problems such as SPIT (spam over internet telephony), SPIM (spam over instant messenger) and Web spam (spam over search engines, which misleads search engines into ranking some pages higher than they deserve).

Index Terms— *taxonomy, spam, Email spam, SPIT, SPIM, Web spam, DDoS attacks*

I. INTRODUCTION

The Internet opens huge opportunities for communication and resource sharing over large distances. Unfortunately, the Internet’s underlying technologies provide only loose coupling between access to resources and accountability for resource consumption. In addition, devices and end-user systems are frequently connected to the Internet with inadequate protection against remotely launched attacks and intrusion attempts. Attackers both seek out and exploit end-system vulnerabilities to gain unauthorised control, or directly attempt to exhaust system and network resources. Port scanning, viruses, worms and Trojan horses are all deployed to break or bypass security, while distributed denial of service (DDoS) techniques overwhelm targeted machines and networks. Completing the problem space, ‘spam’ in various forms brings us excess, uncontrolled and unaccountable resource consumption at the end-user level. Although the term ‘spam’ has been popularised in the context of email communication [1], it is now being more generally used to describe a wide range of mass, unsolicited content distribution. Many studies have emerged in recent years attributing a wide range of end-user costs to spam, usually millions to tens of millions of dollars a year for modest size companies. Pinning down the precise costs is problematic, because it depends so much on the details of each company’s (and person’s) content distribution infrastructure. Nevertheless, spam causes real and substantial consumption of financial and personal resources for recipients and transit

systems.

We believe there is distinct value in defining a scheme for classifying and describing the different ways in which end-user and end-system resources are being uncontrollably consumed by spam in its various guises. A general taxonomy of the problems’ hierarchical relationship will allow developers of countermeasures to understand the different types of related network abuses around their particular problem area. Our taxonomy will help in comprehending the similarities and differences between traditional Email spam and related newer problems such as SPIT (spam over internet telephony), SPIM (spam over instant messenger), Web spam (spam over search engines, which misleads search engines into ranking some pages higher than they deserve) and DDoS attacks.

II. AN OVERVIEW OF SPAM IN ITS VARIOUS FORMS

For this paper we will broadly define a spammer as a person launching abuse of a content distribution system (such as Internet email). A spammer’s primary goal has been to maintain ambiguity about their identity and magnify their ability to distribute content to unsuspecting recipients. *Figure 1* shows the components of a generic content delivery infrastructure that a spammer will use and abuse. We’ve identified transfer agents (TA) as any server that is relaying communication at each network hop, user agents (UAs) as the end-system that originates or consumes the content being delivered, and the ‘Internet’ as the underlying communications fabric interconnecting TAs and UAs. There may be multiple TAs in a path between Sender UA and Recipient UA. Spammers make use of the openness of the Internet content delivery systems to initiate unsolicited transfer of their message towards a Recipient’s UA. To obfuscate their identity to the recipient, spammers will directly inject their content through intermediate TAs or utilise ‘zombies’ – other, unrelated Internet hosts that have been infected or otherwise brought under the spammers control to act as additional Sender UAs for the spammers content. Because the underlying TCP/IP infrastructure lacks strong coupling between access to resources and accountability for resource consumption, large armies of zombies can overwhelm the content being transferred by legitimate Sender UAs.

The specific techniques of spammers have evolved over time. For example, email spammers have evolved from pursuing opportunistic use of open relays and proxies to deliberate infection of innocent machines (through Trojans,

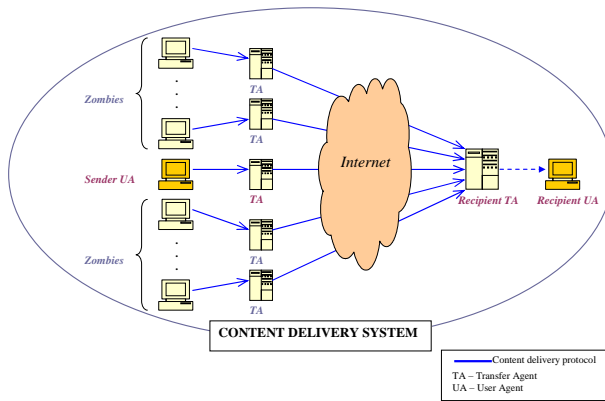


Figure 1 Content delivery infrastructure accounts for excessive resource consumption

worms and viruses) and subsequent creation of large armies of zombie machines (also known as slaves, agents, or daemons). A specific illustration is provided by the behaviour of Sobig.a (and variants Sobig.b to Sobig.f), a virus unleashed on Internet users in 2003 [2]. The virus spread through email attachments that many users accidentally executed, which (in a two-stage process using trojans) installed a key logger and an open proxy program (called Wingate) on the infected machine. This infection turned the host into a zombie UA, allowing spammers to send spam emails with hidden identity. It is likely that the increased prevalence of many viruses, worms and Trojans can be attributed to spammers seeking out new zombies. Other examples include 'Backdoor-Jeen' and 'Backdoor-Guzu' [2]. Internet-based DDoS is also a rising concern. Like spammers, DDoS attackers utilize infected zombie machines as malicious tools to hide the attackers' identity and magnify the damage of their attacks. The relationship between DDoS and Email spam extends beyond the common use of zombie hosts. Email Denial-of-Service (eDoS) and Directory Harvest Attack (DHA) [3] create more intersections between them.

The rapid growth of Internet-based communication services (such as Voice over IP (VoIP), Instant Messaging (IM) and Web page results from Search Engines) made their users as attractive to spammers as email users. Consequently, new terms have emerged - SPIT (spam over IP telephony), SPIM (spam over instant messaging) [4] and Web spam (spam over search engines, which misleads search engines into ranking some pages higher than they deserve) [5].

Unfortunately, Simple Mail Transfer Protocol (SMTP) in email, Session Initiation Protocol (SIP) in VoIP and IM and Web-based search engine all began with minimal authentication and authorisation at each communication hop (between TA to TA and TA to/from UA). In addition, creative methods for infecting unsuspecting end-user hosts create a steady supply of disposable zombies. Thus the incredibly low cost of injecting and transmitting unsolicited content makes spamming an attractive option, whether the spammer's goal is marketing or simply harassment and DoS.

III. PREVIOUS TAXONOMIES

Asami, et al [11] developed their taxonomy for Email

spam using two characteristics: 'delivery schemes' and 'message envelope formats'. The Internet Research Task Force's Anti-spam Research Group (ASRG) has not yet formally published a spam taxonomy, although an informal spam taxonomy was developed by Stumpf on the ASRG discussion list [12]. Stumpf classified emails into four main categories, namely 'private email', 'targeted non-bulk email', 'bulk email' and 'automated messages/answers'. Gyongyi and Garcia-Molina proposed a comprehensive study of Web spam [5] to help the design of anti-spam techniques. Mirkovic and Reiher [7] developed a taxonomy of DDoS attacks and defence mechanisms. They classified different types of DDoS attacks using many criteria, such as: 'Degree of Automation', 'Source Address Validity' and 'Attack Rate Dynamics'. Rosenberg, Jennings and Peterson [6] defined problem areas for spam over the Session Initiation Protocol (SIP) and looked at the solution space of SIP spam in comparison with email. Cerf [4] gave an overview of spam, SPIT, SPIM and their abuse to the Internet resources.

IV. NEW MULTI-TECHNOLOGY TAXONOMIES OF SPAM

The preceding studies are certainly valuable resources for understanding spam in each area. Our study is the first to provide a broad and integrated view of spam across multiple Internet service areas, as well as its relationship with other Internet abuses like DDoS attacks and viruses. In addition to Email spam our taxonomy covers SPIT, SPIM and Web spam - emerging problem areas sharing common traits. Rather than describe all possible details of these problems we focus on characterizing the relationship between various forms of spam, and building a hierarchical structure to facilitate further understanding of factors contributing to the problems. The inclusion of other Internet services vulnerable to bulk, unsolicited communication (such as automatic blog, wiki, and guestbook editing; newsgroup and forum spam; and spam in mobile phones or online games, *et al.*) is a subject for further work.

A. Taxonomy of problem definition

Figure 2 shows our first taxonomy, describing and relating the key message-transmission characteristics of Email, Voice over IP (VoIP), Instant Message (IM) and Search Results (web search). For email, 'VoIP' (voice over IP) and 'IM' (instant message) system we define a 'message' as a communication object that carries content. For web searches we define a 'message' as the relevant content and link (to the search query) [5] of web pages. IM messages are sent atomically in page mode, like regular email messages. Session mode IM requires call establishment signalling before exchanging instant messages. VoIP is similar in that call request signalling occurs before the recipient accepts the call itself (call success) [6]. Our scheme assigns names of the form 'W.x.y.' or 'W.x.y.z' to each path through Figure 2. The first character is A, B, C or D depending on whether we are classifying 'Email', 'VoIP', 'Instant Message' or 'Search result' respectively. The next two or three digits are '1' or '2' depending on which branch of the 'Multiple', 'Recipient Consent' and 'Sender Motivation' categories we take. If the message can be classified after two steps (digits), we do not further divide them.

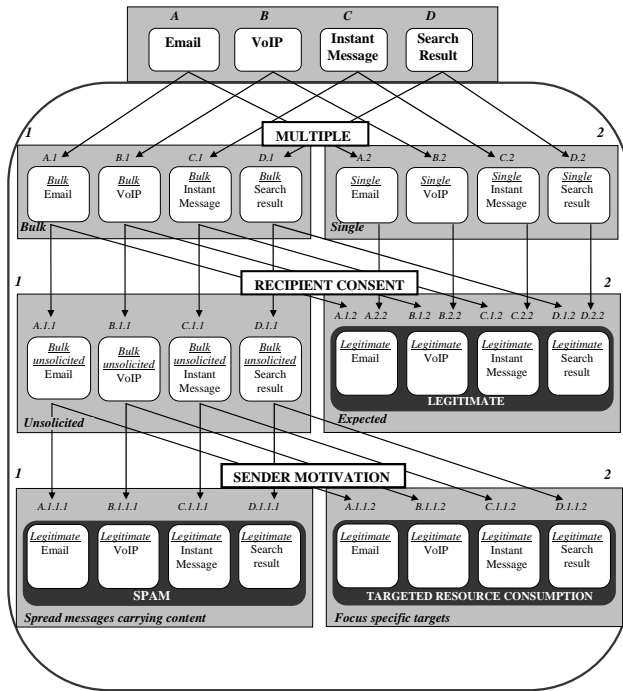


Figure 2 Hierarchical taxonomy of problem definition

The ‘Multiple’ classification reflects the number of messages of the same content sent simultaneously. It is ‘1’ (bulk) if two or more messages are sent or ‘2’ (single) if only one message is sent. In Email, VoIP, and IM systems this relates to the number of messages sent. For Search Results, if the ratio of the messages’ ranking relative to the usefulness of the information provided to a search query is greater than one it is specified as ‘bulk’ (1) otherwise ‘single’ (2).

‘Recipient Consent’ treats messages as unsolicited (1) or expected (2). A message is considered legitimate (not spam) if it is solicited, regardless of whether sent in bulk or single. Search results are considered legitimate if they provide useful information to users whether the results are ranked lower or higher than they are deserved.

Although ‘Sender Motivation’ is hard to practically ascertain it is important to incorporate in the taxonomy. We consider two cases - ‘Spread message carrying content’ (1), where the sender seeks to spread human readable messages as widely as possible, and ‘Focus specific targets’ (2) where the sender’s messages aim at specific targets and the message content is not important. We treat the former case as spam.

Using this scheme, SPIT would be represented as B.1.1.1 – VoIP that is bulk (1), unsolicited (1), and intended to spread message-carrying content (1). Conversely, a normal VoIP call could be B.1.2 or B.2.2. More generally, A.1.1.1, C.1.1.1 and D.1.1.1 are classically ‘spam’ – bulk activities lacking recipient consent and with the goal of spreading targeted content as widely as possible. (For D.1.1.1 these are web pages containing no useful content or links that trick a search engine into giving higher ranking and aim to approach as many users as possible.) A.1.1.2 represents unsolicited emails

sent in large quantities to targeted mail servers or users – this can also include email-Denial-of-Service attacks (eDoS) [3], where excessive resource consumption (of SMTP connections and mail server capacity) causes legitimate email transfers to fail. Similarly, B.1.1.2, C.1.1.2 and D.1.1.2 are all different type of network abuses which aim to cause resource consumption at the recipient targets. These forms of network consumption (classified as ‘Targeted Resource Consumption’) seek to waste the victim’s resources (as opposed to spam where it is intended that message contents are read by human recipients). For example in SPIT, calls are made with human readable content (either by another human or automata) to advertise a product or service whilst messages of B.1.1.2 type in VoIP can be sent by setting up auto-dial, generating random noise (content that is not for human consumption) and then hanging up. Further more, the rest of the messages in our taxonomy A.2.1, B.2.1, C.2.1, D.2.1 (‘Single’ and ‘Unsolicited’) are considered to be socially unexpected messages. The impact of these messages on the recipient is dependent on the social relationship between the sender and the recipient as well as the content being delivered.

B. Taxonomy of main stages in spam sending process

We believe there is commonality between the techniques used by spammers and those used for DDoS attacks (such as ICMP or TCP-SYN attacks, DNS request attacks and email DoS attacks). Figure 3 identifies the four common stages - Magnification, Target, Execution and Hiding Techniques.

Magnification

Magnification refers to multiplication of the number of messages sent. Aside from Web spam this involves recruitment of multiple zombie machines from which spammers/attackers send spam messages or launch attacks. Spammers or attackers then hide behind the identities of each zombie while utilising the zombie’s processing power and network. Zombies are recruited through infection by viruses, worms and Trojans acting on vulnerable hosts (ironically, often themselves distributed as executable attachments to Email spam - one stimulates the growth of the other). Web spam uses several different techniques to boost content relevance, such as ‘dumping’ a large number of unrelated terms to be relevant to many different queries [5]. Web spam may also create specific link structures designed to increase the link ranking of their pages (such as a ‘spam farm’ – large collections of inter-linked pages to boost some page ranks).

Target

In this stage we classify the techniques used to identify targets or legitimate addresses identifying real users. The specific address formats differ between Email spam, SPIT and SPIM but the basic principles are the same. Spammers can try addresses randomly, find addresses at public message boards, news groups, and mailing lists, or buy address lists from spam dealers who collect and sell users’ address. Directory Harvest Attack (DHA) [3] against mail servers can also be launched to obtain valid email addresses for an email domain. With DDoS attacks or Web spam the target is always ‘known’. DDoS attacks on applications, hosts, resources, networks and infrastructure [7] use pre-defined targets before the attack is launched. Web spam targets search engines whose searching

algorithms and related searching and ranking methods are ‘known targets’ to spammers.

Execution

Execution describes the spammer-triggered processes that actually instantiate content transfer - launching DDoS attacks, sending spam emails, SPIT and SPIM. For Web spam this is ‘not applicable’ as the search engines initiate searches in response to search queries regardless of the presence of spam pages. DDoS attackers launch different types of attacks (IMCP, TCP SYN, DNS request, email, *et al.*) towards the targets in order to cripple the targets or prevent legitimate access to the targets. Emails are sent using the simple mail transfer protocol (SMTP). Email spammers may automate this process with commercial spam generators/managers to rapidly send, and track the sending of, a large number of emails. SPIT utilises the session initiation protocol (SIP), which is also emerging as a generic signalling standard for IM systems [6]. VoIP calls are established by sending SIP ‘Invite Request’ messages to a recipient, who must positively acknowledge the request before call content is transferred. IM in session mode has a similar pattern of call request - recipient answer. IM in page mode is analogous to the email system, in which each IM is sent in a separate message. To automate and magnify the sending process SPIT and SPIM are usually sent by automata (commonly known as bots).

If the magnification stage involves zombies the execution stage relies on those zombies. DDoS attackers and email, SPIT and SPIM spammers can utilise specific, unadvertised TCP or UDP ports with which to communicate with their zombies (to the zombie software previously installed in the magnification step). Infected zombies may also ‘listen’ for control messages on specific Internet Chat Program (IRC) channels, making it hard to distinguish from legitimate IRC traffic.

Hiding techniques:

Spammers and attackers try to hide their identity from everyone, and hide their message content (or attack intentions) until it reaches the intended recipient. The ‘Hiding’ stage covers both techniques used to conceal identities and mechanisms to obscure spam content of messages.

Zombies provide a level of indirection that hides the controller or source of spam or DDoS attack – the recipients see the zombie’s IP address as the source, and there’s frequently no inter-ISP mechanism to trace back through a zombie to the real source. Address spoofing hides a message’s source by inserting false ‘source’ information in the message. For example a fake IP source address in packets sent as part of a uni-directional DDoS attack, a fake ‘From’ address in a spam email, or a misleading website address embedded in an HTML-encoded email. The latter two examples are often combined in email phishing [8], where false source email addresses and carefully crafted HTML-emails lead people to reveal their legitimate online banking details to a fake ‘bank website’. A number of techniques, such as Microsoft’s ‘Sender ID Framework’ [9] and Yahoo’s ‘Domain Keys Identified Mail’ [10], have been proposed to authenticate an email’s true source and thus minimise incidences of phishing. (SIP-based content delivery systems are less susceptible to address spoofing - a SIP sending domain authenticates its user

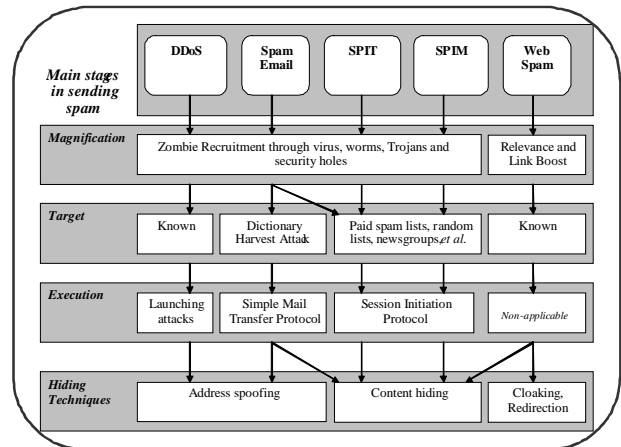


Figure 3 Taxonomy of main stages in spam sending processes

and includes the identity of the user and its own signature to be verified by the receiving domain [6].)

Content hiding tries to side-step anti-spam tools that perform textual or statistical content analysis. Most examples exist in the email space today, although one can easily imagine analogous techniques appearing in the SPIM (and perhaps SPIT) space. For example, email spammers may split key spam words by inserting a letter or substituting a character with a different one but of similar look (e.g. number ‘0’ for letter ‘O’). Web spam uses various forms of content hiding to ensure that the search terms used to push up a page’s rankings in the search engine do not appear when user’s actually visit the spammer’s site. For example, pages may hide additional keywords using text with the same colour as the background. Alternatively the web spammer can use cloaking (a page with spam content is returned to users while a different page is returned to the web crawler used by the search engine) or redirection (the page with high ranking is immediately redirected to a page with spam content).

V. THE VALUES OF OUR TAXONOMIES AND OUR PROPOSAL TO SPAM MITIGATION

A. The values of our taxonomies

Firstly our taxonomy of spam definition is semantically useful in terms of describing the problem. It shows the three important classification features and how a message in a content delivery system should be defined as spam. It suggests viewing spam from an integrated angle to understand the difference and similarity of spam across different areas.

Secondly our taxonomy of the spam sending process illustrates techniques involved in the main stages of spam distribution and DDoS attacks. Understanding the common traits of spam and DDoS attack techniques in different areas allows network administrators to develop useful Intrusion Detection Signatures and anti-spam methods to protect their network from the excessive resource theft.

Finally, is it is an important part of our taxonomy to explicitly recognise ‘Sender Motivation’ as a key component in spam classification process. RFC 2505 acknowledges only ‘Mass’ and ‘Unsolicited’ (in our terms, ‘Multiple’ and

‘Recipient Consent’) as characteristics of Email spam, neglecting the reality of sender motivation. Although not officially recognised, existing anti-spam schemes implicitly attempt to infer ‘sender motivation’ as part of their spam identification process. Anti-spam techniques such as black-listing and white-listing assert something about a sender’s motivation based on their past history (whatever events caused them to be on the black- or white-list). However, the link between past history and the sender’s current motivation may be weak or entirely broken unless the sender’s behaviour is continually monitored. Content classification tries to infer a sender’s intent by ‘looking for spam’ in the received messages. Unfortunately, there are note-worthy rates of false-positives (legitimate emails falsely classified as spam) and false-negatives (spam falsely treated as legitimate emails).

B. *The problem facing current anti-spam techniques and our proposal to the question of Sender Motivation*

After recognising that simply classifying spam based on the two basic characters (‘Mass’ and ‘Unsolicited’) was inadequate, many have tried to incorporate the Sender Motivation into their technique. Generally there are two fundamental solutions– ask sources to prove they are human beings interested in legitimate communication, or require sources to prove their willingness to actually expend resources to send a message. The former is exemplified by challenge-response (C/R) schemes - an unknown sender is challenged by an automated ‘request email’, and unless the senders replies correctly to the challenge message the original email will not be delivered to the recipient. C/R fails where two communication ends are unknown to each other and both are set to use C/R mechanisms; or when the spammers know the C/R rule and set their spam robots to bypass this rule. Microsoft’s computational payment (puzzle) approach [13] is an example of an alternative, ‘economics-based’ scheme. Email senders must solve cryptographic puzzles set by the recipient. Puzzles are designed to consume so much time and processing resources at the source that a spammer finds it infeasible to send thousands of emails but a ‘legitimate’ sender can tolerate the cost. However, spammers can distribute this computational load across their zombies and reduce the incremental cost to their operation. Moreover, when emails from thousands of zombies converge on one recipient - the recipient may suffer a form of eDoS. This is an issue faced by any scheme that presumes a willingness to incur ‘computational cost to send’ implies something about the sender’s motivation.

Our suggestion for Sender Motivation is to constantly re-rate the source using content analysis that dynamically updates a short-term black-list. When a threshold is reached, the black-list triggers IP packet-level rate limiting creating additional network-layer resource consumption at the source [14]. The overall cost of sending Email spam increases (for spammers), yet the impact of false-positives (on non-spammers) is reduced (falsely classified email still eventually gets through). Since the sender’s recent, content-analysed behaviour heavily influences a recipient’s local black-list a sender can

rehabilitate themselves automatically over time, without waiting for external black-list maintenance to clear their name.

VI. CONCLUSION

We investigate the relationship between spam in email and other content delivery systems. We propose a multi-technology taxonomy covering the problem definition, and a second taxonomy covering various methods of sending spam. Our first taxonomy shows how messages in different content delivery systems may be categorised as spam or not spam. We suggest ‘Sender Motivation’ as an important factor to consider in classifying spam and deploying anti-spam techniques. Our second taxonomy covers the main stages in sending spam and provides a more detailed view of the techniques involved in injecting, multiplying and propagating unsolicited content across the Internet. We identify similarities and differences between the mechanisms used to instantiate Email spam, SPIT, SPIM and Web spam. We also examine features shared by spam and DDoS attack mechanisms, and illustrate how they contribute to the consumption of the Internet resources.

REFERENCES

- [1] T.Bass, G.Watt, “A simple framework for filtering queued SMTP mail (cyberwar countermeasures)”, *Proc of 1997 IEEE Military Communications Conference (MILCOM 1997)*, California, USA, Nov 1997
- [2] E. Levy, “The making of a spam zombie army. Dissecting the Sobig worms”, *IEEE Security & Privacy Magazine*, Vol. 1 (4), pp. 58-59, Jul 2003
- [3] Tumbleweed, “Secure Email Relay and Defense Against Dark Traffic”, <http://www.tumbleweed.com/> (as of Aug 2006)
- [4] V.G. Cerf, “Spam, Spim and Spit”, *Communications of the ACM*, Vol 48 (4), pp 39-43, Apr 2005
- [5] Z. Gyongyi, H. Garcia-Molina, “Web Spam Taxonomy”, *Proc. of the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb 2005)*, May 2005, Chiba, Japan
- [6] J. Rosenberg, C. Jennings, J. Peterson, “The Session Initiation Protocol (SIP) and Spam”, *Internet Draft, Internet Engineering Task Force*, Jul 2005
- [7] J. Mirkovic, P. Reiher, “Taxonomy of DDoS Attack and DDoS Defense Mechanisms”, *ACM SIGCOMM Computer Communication Review*, Vol. 34 (2), pp. 39 – 53, Apr 2004
- [8] C.E. Drake, J. J. Oliver, E. J. Koontz, “Anatomy of a Phishing Email”, *Proc. of the First Conference on Email and Anti-Spam (CEAS 2004)*, Mountain View, CA, USA, Jul 2004
- [9] “Sender ID Framework Overview”, Feb 2005, <http://www.microsoft.com/mscorp/safety/technologies/senderid/overview.msp>
- [10] M. Delany, “Domain-based Email Authentication Using Public-Keys Advertised in the DNS (DomainKeys)”, *Internet Draft, Internet Engineering Task Force*, Sep 2005
- [11] T. Asami, T. Kikuchi, K. Rikitake, H. Nagata, T. Hamai, Y. Hatori, “A Taxonomy of Spam and a Protection Method for Enterprise Networks”, *Proc. of the 16th IEEE International Conference on Information Networking (ICOIN 2002)*, Cheju Island, Korea, Jan 2002
- [12] Markus Stumpf, “Taxonomy - Classification of messages”, ASRG Discussion List, Oct 2003, <http://www1.ietf.org/mail-archive/web/asrg/current/msg07651.html>
- [13] The Penny Black Project, Microsoft Research, <http://research.microsoft.com/research/sv/PennyBlack/> (as of Aug 2006)
- [14] M. Tran, G.Armitage, "Evaluating The Use of Spam-triggered TCP/IP Rate Control To Protect SMTP Servers," *Proc. of the Australian Telecommunications Networks & Applications Conference 2004 (ATNAC2004)*, Sydney, Australia, December 8-10, 2004