An Overview of Spam Phenomenon; and the Key Findings of a Survey for Spam in Greece

Anastasios A. Pallas[1], **Charalampos Z. Patrikakis**[2]

[1]*School of Computing,*
University of Paisley,
PA1 2BE, Scotland, tpallas@sch.gr

[2] *TEI Peiraeus, Automation Deptm.*
Thivon 250, Aigaleo,
Athens, Greece

Abstract

Spam has been involved in the every day life of e-mail users. They actually recognized it when they get annoyed with advertisements, bulk e-mail messages or in general all the unwanted e-mails in their mailbox. Recently, spam phenomenon has been also found to "speak" the Greek language; and some international flavors to shifting from annoying to threatening Internet users, which in effect constitute a major security and fraud issue over Internet. In the current paper, apart from introducing spam problem and providing an overview of its current countermeasures (anti-spamming), we present the most significant key finding from our survey for the spam phenomenon in Greece. Our areas of interest cover some of the major spam issues such as: awareness, impact, costs, legislation and anti-spamming.

Introducing Spam

The actual provenance of the word "spam" is derived from the words "spiced ham" (first entry in 1937) for describing a tinned meat made mainly from ham by Hormel Foods Corporation. In the electronic world, "spam" has not an official definition, thus some people consider advertisement via e-mails as "spam", and others consider "spam" as just all unwanted e-mails in their mailbox even though they are coming from their mammy.

Actually, there are two rather official definitions for "spam": Unsolicited Commercial E-mail (UCE) and Unsolicited Bulk E-mail (UBE). The term UCE is most frequently used in United States and widely used into its most important spam law formally called **CAN-SPAM Act of 2003** [3]. UCE is explicitly used for commercial messages or in other words advertisements, such as messages with pornographic content, marketing of illegal software, etc. Although, the definition of UCE (along with CAN-SPAM Act of 2003) is very important, it excludes messages that can be "spam" and they are clearly not commercial in nature, for example political, frauds, malicious messages and many others, which are covered by UBE definition. General speaking, UBE is more broaden definition of spam covering messages sent to a large number of recipients who have not opt-in (i.e. give prior consent in written or electronic format) or have opt-out (i.e. asked somehow to unsubscribe). Besides, US law has a major drawback as it allows e-marketers to legitimately send messages as long as a recipient has not opts-out (i.e. opt-out approach to spam problem). In contrast within Europe Union, spam is prohibited by the **European Directive for e-Privacy 2002/58/EC**, which clearly requires that commercial communication should not be allowed without recipients' prior consent, meaning that recipients have to prior opt-in in order to legitimately receive commercial messages (i.e. opt-in approach to spam problem) [4].

Anti-Spamming

Spam is meant to be one of the most critical problems upon Internet, nowadays. First of all, it wastes Internet users' precious time. It is also contributes to the misuse of Internet bandwidth, computers' processing power and storage capacity worldwide. Furthermore, there are also many hidden and rather difficult to be measured effects due to spam, such as the lost of legitimate e-mails – namely False Positives (FP) effect–, the misleading of Internet consumers, exposure to unethical content for children, electronic frauds, etc. As a matter of fact, there have been also deployed a number of countermeasures, which are trying to moderate spam phenomenon (namely anti-spamming). For clarity reasons, we are going to present here a rather conceptual view of these approaches and not drill down to technical analysis. **In general, there are three primarily not overlapped approaches: legal, social and technical** [10].

Providing a more detail analysis of the efforts held **in the technology-based approach, we can further distinguish countermeasures based on: i) where they reside and ii) how they react against spammer's techniques**. In the first case, we can distinguish anti-spamming efforts that reside either to the side of the server or in other words to the Mail Transport Agent (MTA) of an e-mail service provider (i.e. namely mailer-based or server-based) or to those that run to the user's computer (i.e. namely user-based or client-based). In the second case, we can depict anti-spamming efforts into **complementary approaches** versus spammers' methodologies [5]:

- **Protect e-mail addresses**: Spammers are trying to harvest e-mail addresses by using a number of methods, such as dictionary attacks to mailer and collect them from web sites. Preventing an e-mail address to be listed from spammers is mainly a set of directives and techniques that users can adopt. For instance, a rather effective technique has been proven to be the masquerading of e-mail addresses when they are going to be used in web sites, forums, etc by replacing for example the "at" symbol (@) or the "dot" (.) or by using more complicated methods [7][8].
- **Prevent spam from being sent**: It is very common for spammers to use compromised/hijacked computers (also called as zombie) all around the world in order to send unsolicited messages. Preventing spam from being sent involves not only the regular checks of computers' security holes, but also blocking of SMTP and proxy relays.
- **Block spam to be delivered**: There are a number of different techniques here that are trying to identify spam messages before actually reach our mailbox. Some of them are checking sender authentication such as Sender Policy Framework (SPF) and blocking exploited sources of known spammer based on IP or DNS (i.e. DNSBLs[1]) [16]. Usually these methods have been widely adopted by Internet Services providers in order to refuse delivery of spam messages.
- **Identify and separate spam after delivery**: Spammers are continually inventing new tricks in order to spoof anti-spamming filters. Identifying spam based on content analysis of the messages has been proven to be very difficult indeed. However, two methods have been proven here to be more effective: i) the analysis that is based on the targeted link (i.e. URI[2] analysis and SURBLs[3]) and ii) Bayesian filters that are relying on adaptive filtering algorithms.

As we already mention, the above technology-based techniques are basically complementary and not primarily overlapped. Thus, **these techniques are common not to be used alone, but rather to be invoked into the so called "heuristics" or rule-based scoring systems**. As a matter of fact, heuristics is not so much a unique method rather a framework for combining various tests (such those we mentioned above) and assigning relative scores to their results; the summary of the resulting scores

[1] Domain Name System Blacklists or Blocklists
[2] Uniform Resource Identifier
[3] Spam URI Real-time Block-lists

is finally compared to a preset threshold. If a message' score exceeds the threshold then it is assumed that it is spam or if it is below the threshold then it is "ham" (i.e. not spam) [12][13].

A Survey for Spam Phenomenon in Greece

The scope of our survey was to investigate the spam phenomenon in Greece, through the acquisition of feedback from the major e-mail players, such as Greek Internet service providers (ISPs), independent e-mail service providers, and academic institutions. The data collection techniques invoked the study of spam-related documentation, questionnaires and interviews with experts. Our survey data was collected via interviews with "experts" and the procedure was guided by a specialised questionnaire consisted with quantitative and qualitative data. In our case, expert consisted by IT directors, e-mail administrators, system administrators and other IT technical staff that have strong relationship in the administration of their e-mail systems at strategic or technical level.

The key issues investigated in our survey were: i) spam awareness, ii) impact of spam, iii) costs related to spam, iv) spam legislation, v) anti-spamming, vi) Greek spamming and vii) the future of spam. Our questionnaire comprised of a total of sixty seven (67) questions that most of them were in closed-end format; and in the majority of the questions answerers were provided with the ability not to answer the question either due to unwillingness (i.e. "I prefer not to answer") or due to insufficiency of knowledge (i.e. "I don't know"). Apart from the questionnaire, the discussion with the e-mail experts was indeed beneficial, covering trends on spamming and anti-spamming, as well as, feedback on how they currently handling the phenomenon and what are their opinion about spam.

The survey took place in 2005 with the participation of experts from the major Greek ISPs, e-mail providers and educational institutions. It is important to say that our sample covered a significant portion of the population of the major players that exist nowadays in Greece; and with regards to the people participated, the majority of those were e-mail experts employed in educational institution and the rest were working for ISPs and independent e-mail hosting providers. Furthermore, **the majority of experts interviewed were system administrators and e-mail administrators**, some of them general directors and a small portion were technical personnel.[9]

Key Findings and Recommendations

The key findings derived from our survey can be summarized to the followings:

- As it was expected **defining spam seems to be a little confusing**; however, most dominant opinions are identifying spam as UCE and UBE.
- Spam is rather a very important issue to our sample of organizations.
- The percentage of e-mails identified as spam in Greece seems to be quite similar to the global spam statistics [11][12]. This indicates that **spam has analogous impact to Greece as worldwide**; and as a matter of fact, **Greeks should increase their concern about the spam phenomenon irrespectively to the difference of language**.
- Almost all providers have been established some low-cost (mainly software-based) countermeasures to moderate spam effects, as well as, to eliminate users' complains.
- Regarding spam legislation, it is essential that experts already recognize that spam phenomenon is not explicitly dedicated to the e-mail communication and they would like legislation to be rather sufficient; and why not stand alone.
- **Our sample believe that the regulation of spam passes through legislation, penalties and an appropriate national agency, which with the assistance of ISPs should be able to bring actions against spammers**. Major players seem also willing to provide necessary information, but they are skeptical to provide details regarding their customers.

- **The majority of providers use for anti-spamming the well-known DNSBLs and a significant portion has also enable heuristic techniques and custom filters to the server-end**.
- Providers seem also to be quite satisfied with their tools, but they are concerned about filtering accuracy and the involvement of users to the process.
- A major problem that came to the surface is the manipulation of the identified spam, where our survey identified some misuse. Unfortunately not all messages identified as spam are really spam. By either moving these messages to separate folder or deleting them, we increase the change of missing important legitimate messages.
- Finally, almost half of experts believe that spam in Greek language are insignificant compared to the English-based and that it is quite rare; and for a portion of them also believe that it is difficult to be identified by anti-spam tools. Furthermore, half of the experts seem also worrying about the impact of spam to the spread of Internet in Greece.
- Finally, experts' forecasts are unfortunately discouraged for both spam phenomenons worldwide, as well as, for spam in Greek language.

We believe that **spam phenomenon tends to be social rather than technological, and needs a unified approach and cross-border cooperation** in order to be effectively reduced. Spammers are meant to take advantage of any broadband medium available and not only e-mail communication. Spam in mobile communication is actually expected to be highly increased into the following years and become a critical issue comparative with spam in e-mail communication [2]. Much the same, spam is expected to be spread into Internet Telephony and Voice over IP (VoIP) under the name "SPIT" [4], as well as, into Internet Instance Messaging (IM) (eg. ICQ and MSN Messager) under the name "SPIM"[5]. [10]

It is also amazing that recently **spam has been also found to "speak" Greek**. Indeed, during our experiments with the use of spam traps[6] that held from 21st March 2005 to 10th October 2005 (204 days), 10 spam messages found to be in Greek out of the total 125 spam messages, representing the 8%. Even though, Greek spamming is still rather small in contrast to the English-based, we should take proper actions to eliminate it before is too late. On the same hand, **some international flavors of spam are shifting from annoying to threatening** Internet users' society. As a matter of fact, the amount of electronic frauds traveling though spam under the name "Phishing"[7] and "Pharming" has dramatically increased the last years [1][6]. With the term phishing, we actually mean the exposure of Internet users' sensitive information such as credit card numbers, bank accounts and access passwords; and with the term pharming, we mean the decoy web sites where intruders redirect users in order to harvest their data [14][15].

Under these circumstances, we believe that **national bodies should take immediate actions against spam phenomenon**, set up or appoint an appropriate agency/operator that will lead into activities against spammers and their sponsors. On the same hand, with spam tending to constitute a major Internet security and e-fraud problem **Internet Service Providers (ISPs) should also take actions by providing assistance to enforcement agencies along with undertaking of users' education**. We believe that ISPs should offer continuous and free guidance to users against spam, which should include what they should be aware, how to avoid spam, and how they should react.

[4] Spam over Internet Telephony
[5] Spam in Instant Messaging
[6] A decoy e-mail address that is explicitly used for collecting spam is called spam honey pot or spam trap.
[7] Phishing is derived from the words: Password Harvesting and fISHING.

Acknowledgements

Authors would like to mention that the survey was part of the MSc thesis of Mr. Anastasios Pallas under the title "*The spam phenomenon: techniques, countermeasures; and its distinctiveness in Greece*"; and they would like to thank the people participated into this survey for spam, for their assistance and useful feedbacks.

References

[1] Anti-Phishing Working Group (APWG). Phishing Activity Trends Report. APWG, September 2005 (www.antiphishing.org/apwg_phishing_activity_report_sept_05.pdf)
[2] Brodt, Torsten and Janos Hee. "Insights into Mobile Spam: World's First Collaborative Empirical Study". University of St. Gallen, Switzerland, 2005.
[3] CAN-SPAM Act of 2003, Public Law 108-187, 108th Cong., 1st sess., 16 December 2003 (www.spamlaws.com/pdf/pl108-187.pdf).
[4] Directive 2002/58/EC. Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector. 2002 O.J. (L 201) 37 (www.spamlaws.com/docs/2002-58-ec.pdf).
[5] Koh, Eung Lyeol. "Anti-Spam Toolkit: Existing and Emerging Technical Measures against Spam." Slide presentation. Spam Response Team, Korea Information Security Agency, 8 September 2004.
[6] Matthew P., (2005). "'Pharmaing' Is Latest Internet Security Threat". Inc.com, 20 June 2005 (www.inc.com/criticalnews/articles/200506/pharming.html)
[7] Morris, John, (2003). "Best Practices for End Users". Presentation to Internet Engineering Task Force 56, San Francisco, California, USA, March 2003 (www.ietf.org/proceedings/03mar/slides/asrg-3/index.html).
[8] Pallas A., (2005). "Hide your e-mail from spammers" (www.no-spam.gr/mustknow.htm)
[9] Pallas A. and Patrikakis C., (2006). "The Spam Phenomenon in Greece, Countermeasures and Future Trends". Proceedings of the 2nd Conference on Electronic Democracy: Challenges of the Digital Era. Athens, Greece, March 16-17, 2006.
[10] Patrikakis C. and Pallas A., (2006). "Are we ready to face next-generation spam?". Cutter IT Journal. January 2006 Vol. 19, No 1. ISSN: 1522-7383.
[11] Symantec Corporation (2005). "Symantec Spam Statistics - Monthly report analyzing Symantec's probe network data to identify trends". May 2005 (www.symantec.com/region/reg_ap/promo/brightmail/docs/May2005SpamStats.pdf).
[12] The Apache SpamAssassin Project (2004). "Spam - What is spam?" SpamAssassin Wiki (wiki.apache.org/spamassassin/Spam).
[13] The Spamhaus Project, (2006). "Effective Spam Filtering" (www.spamhaus.org/effective_filtering.html)
[14] Wikipedia Web site (http://en.wikipedia.org/wiki/Pharming).
[15] Wikipedia Web site (http://en.wikipedia.org/wiki/Phishing).
[16] Wong M. and Schlitt W., (2005). "Sender Policy Framework (SPF) for Authorizing Use of Domains in E-MAIL". Network Working Group Internet Draft, IETF (www.ietf.org/internet-drafts/draft-schlitt-spf-classic-02.txt).