IGI PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

ITJ 4725

This paper appears in the publication, International Journal of Digital Crime and Forensics, Volume 1, Issue 2
edited by **Chang-Tsun Li © 2009, IGI Global**

Methods to Identify Spammers

Tobias Eggendorfer, Universität der Bundeswehr München, Germany

ABSTRACT

Unsolicited commercial email has become a major threat for email communication. Although the degree of sophistication of spam filters has increased over time, such filters still produce high rates of false positives and false negatives, thereby reducing the reliability of email and introducing communication risks on their own. Due to more and more complex filtering methods implemented, the hardware requirements for mail servers are increasing to avoid the risk of denial of service situations. Therefore, some authors point out that mail filtering has reached its limits and ask for more preventive solutions to fight spam. One way to prevent email abuse would be to significantly increase the risk of a spammer being sued for damage compensation or, if legislation permits, for criminal offence. This approach believes in an assessment of risk and expected revenue by the offender. But by hiding their real identity, spammers are very successful in evading prosecution. This paper discusses several methods to identify spammers and analyses under which circumstances those methods might be valid evidence in court. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: *Address Trading; Forensics; Identification; Spam*

INTRODUCTION

Although not anticipated by the founders of the Internet, email has become one of the most accepted and widely used applications of the Internet. But with an ever increasing proportion of unwanted email, users are slowly starting to think about switching to other means of communication. Some use instant messaging instead, others are returning to the fax, despite being more expensive and less convenient.

Spam is not only an inconvenience, it also has high associated costs, including the wasted work time reading unsolicited emails, the investments in hardware and software needed and the costs associated with maintaining spam filtering systems. Additional costs are also incurred for providing the necessary bandwidth and extra hardware to handle the flood of emails.

Although the definition of spam seems to vary, with some authors restricting it to unsolicited commercial email and others broadening it up to any unsolicited bulk email, including mass emails sent to distribute viruses, worms and Trojans, hoaxes and even chain letters, they share the observation that spam makes up for the vast majority of all emails sent worldwide, be it more than 80% in July 2007 according to spam-o-meter (2007) or even more than 97%, as claimed by T-Online, one of Germany's biggest email providers (Kuri, 2006).

If providers were able to filter all of this unsolicited email automatically and with a zero error rate, email users would not care about spam, because they would not receive any. In an ideal world, where perfect filters exist, spammers would even have to discontinue their business because their earnings would drop to zero. The spam problem would then be resolved.

Unfortunately, spam filters only offer more or less accurate heuristics to help sorting spam and ham, as legitimate email is often called. Therefore, reality is far from this perfect world.

False Positives and Negatives

Recent surveys (Eggendorfer, 2007b; Schulz, 2006; Hosbach, 2006) found that false positives rates of spam filters might be as high as 18% and false negatives easily reach 20%. Although false negatives, i.e. spam not marked as spam, are annoying to the user, false positives are of far greater concern: in a business environment a false positive might have been a customer ordering a product. Failing to notice this message due to an overacting spam filter might not only mean a loss in sales but also liability for not delivering the requested products, thereby increasing the potential financial losses from a false positive by orders of magnitude (Heinlein, 2007).

Although seldom considered to be so, spam filters therefore might be one of the risks associated to spam, even if they help to cure some of the symptoms of the spam epidemic.

On the other hand, false negatives, often considered to be only annoying, but not a security risk, also introduce a new risk: The human false positive. The more unwanted messages actually make it to the end user's mailbox, the more messages the user has to filter manually. To do so, the user often only relies on the sender address and the subject line. Considering how many other factors a spam filter uses to identify a legitimate message, the lack of precision is obvious, even if one believes that human intelligence is likely to be superior to simple computer based heuristics and might find more evidence for a message being spam than a computer does by taking more factors into account.

However good the human brain might be, the sheer quantity might lead to users accidentally and unintentionally marking one message too many for deletion or as spam. Those misidentifications by the user are "human false positives" and are at least as troublesome as machine false positives. Human false positives are unpredictable in their nature and cannot be avoided by technical means, except with better spam filters or methods that decrease the total amount of spam. Yet, the more aggressive a filter is, the more likely are machine false positives (Eggendorfer, 2007b). This is a *circulus virtuosus*.

Security Considerations

Also, spam filtering increases the risk of security leaks on an SMTP server: the more complex filters are (some even implement OCR to identify image spam) the more computing power they consume; the more power they consume, the higher are the requirements on the mail server's hardware with constant processing time per message, or the longer the mail processing time becomes on unmodified hardware. With each and every message taking longer to be processed, the mail server will be able to handle less requests per second. This again increases the risk of a denial of service attack on the mail server. Apa (2003), Frei (2004) and Schüler (2004) provide anecdotal reports of this being more than just a theoretical idea. The author notes he recently had to upgrade his mail processing system to more powerful hardware due to the heavy load his mail and virus filtering generated.

Not only is there a risk of denial of service attacks, but also does each additional line of code on any system mean an increased risk of this programme containing one or more bugs. Bugs might introduce security leaks, which might be exploitable remotely. Those security leaks might be used to attack the mail system and compromise it. Obvious risks are abusing