Privacy and Spam: Empirical Studies of Unsolicited Commercial E-Mail

Andreas Jacobsson & Bengt Carlsson
Department of Software Engineering and Computer Science
Blekinge Institute of Technology, PO Box 520, S-372 25 Ronneby, SWEDEN
{andreas.jacobsson;bengt.carlsson}@bth.se

Abstract. In our society, the Internet becomes more and more indispensable, and the issue of personal information between consumers and businesses is recognised as a critically important one in building a secure and efficient (social) system on the Internet. When it comes to handling personal information, consumers generally want their privacy to be protected, but businesses need reliable personal information and an access channel to consumers for e-commerce. Undoubtedly these demands must be satisfied to establish sound e-commerce. However, with the technologies available today it is reasonably easy for companies to gather information about consumers in order to make personalised offers through e-mail. There is a fine line between collecting personal information to make customised offers that customers regards as useful information and what is an intrusion to personal privacy. This paper discusses how consumer privacy is affected by unsolicited e-mails sent with a commercial purpose (spam). We found that, albeit most of the investigated web sites behaved well, a small fraction generated a large number of spam.

1 Introduction

Today, many businesses provide personalised services and offers to their customers by utilising customer preference information and/or behavioural information. Various web sites can collect each web user's browsing log and display personalised pages for them by using, for example, cookies. These customised services certainly make life more convenient. Therefore, it is reasonable to allow a business to provide such services. However, whether a business is successful or not, often depends on having more information about consumers than the competitors do. This means that a business must gather as much information about consumers as possible, which on the other hand increases the possibility of unintentional, or intentional, infringement of consumers' privacy. Continued abuse of consumers' privacy makes them very uneasy about sharing their personal information with businesses, and possibly also sceptical to the idea of e-commerce. Human society on the Internet will only thrive if the privacy rights of individuals are balanced with the benefits associated with the flow of personal information [4].

Here, we use the definition first proposed by Samuel D. Warren and Louis D. Brandeis in their article "The Right to Privacy" [5] and define privacy as "*the right to be let alone*". The advantage of this definition is that it is widely accepted in society, and thus easily can be adopted to the Internet setting. The extraction of the definition is that users can specify what information should be disclosed to whom, when it should be disclosed, and for what purpose. Also, that they are guaranteed that the information will be treated

accordingly. In effect, this correlates well with the most commonly used definition of privacy by Alan Westin:

> "Privacy is the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others." [5]

In general, the explicit demands from consumers and businesses regarding commerce based on the customers' personal information do not always fit well with having the right to be let alone. Consumers want to be able to control their privacy and still get the best personalised services available. Businesses, on the other hand, need reliable personal information about customers, and also an access channel to bring the best service possible to the appropriate consumer. From this discussion we can extract three requirements for sound e-commerce based on personal information:

1. **Privacy control**: Consumers should be provided with a means to decide what, when, and for what purpose their personal information is used.
2. **Data reliability**: Businesses should be provided with reliable consumer information.
3. **Consumer accessibility**: Businesses should be provided with a means to access targeted consumers directly.

As can be seen later in this paper, these requirements are not fully met when it comes to e-businesses informing users about offers. In effect, this is a problem that needs cooperation from both customers and companies. In the next sections we will conduct a slightly empirical discussion regarding privacy requirements and unsolicited commercial e-mails. We conclude this paper with some final remarks.

2 E-Mail Marketing

2.1 The Spam Situation Today

Unsolicited commercial e-mails, also known as spam or junk e-mail, has increased dramatically in number over the last years. It has become one of the most used marketing tools for the Internet. In the beginning of 2002 estimations showed that one out of twelve e-mails fit the description for spam. During that year the number of spam rose and reached an average frequency where one out of three e-mails were junk e-mails. An assessment made this year by Ferris Research estimated that it takes the average Internet user about 4.4 seconds to handle a spam, and also that approximately 20 billions such e-mails are sent every day from data bases holding up to 200 million e-mail addresses [1]. The accumulated time for handling spam approaches 25 million hours per day. On a personal level, at least a few minutes daily goes by for deleting unsolicited e-mails.

2.2 The Contents of Spam

In two experiments, performed during the spring of 2003 at the Blekinge Institute of Technology, the occurrence of spam, and their impact on personal privacy were investigated.

In a master thesis experiment, 30 different well-known and highly trafficked web sites were selected [2]. The web sites were equally selected from the United States and the European Union with two fictious individuals, Adam and Bill, representing average In-

ternet users from the US and EU respectively. By signing up the web sites mailing lists the intention was to examine if personal information was spread to third parties and/or if it generated spam. During each session Adam and Bill registered numbered e-mail addresses, i.e., adam1@ourdomain.se, adam2@ourdomain.se, etc., to clarify which e-mail address was added to which web site. It turned out that only one, Music.com, out of 30 web sites visited generated spam. Furthermore it was only Adam, the pretended American visitor, who received spam.

By the second day of the test Adam got his first spam and after that it increased during the forthcoming two weeks. In all 468 spams were received over a five-week period analysed (see Figure 1). The first spam offered an insecure Gold Card and was sent from Arbango.com. The company claimed that Adam had requested to receive special promotional messages from Arbango.com, however such a request was never made. These kinds of claims have also occurred frequently in other spam received as well as statements that the e-mail address was passed to the spammer by an alleged friend. In some of the e-mails Adam were greeted by his name (Adam Smith) though the e-mail address did not reveal his last name. So, there was no connection to the original Music.com site when examining the contents of received spam. In all, spam from 15 different categories were received, see Figure 2 for results.

In a second experiment we investigated if unsubscribing to spam e-mails generated new spam. In all 219 spam were investigated and 182 of them allowed the user to unsubscribe. By transporting the spam messages to a newly configured e-mail account with a "clean" environment it was possible to investigate the impact of unsubscribing to spam. During a four-week period we did not receive a single spam in return.

In the first experiment the categories that have generated the most spam were *free offers* and financial advertisements such as offers to *loan* and *make money*. This finding somewhat correlates with what Cranor and LaMacchia found in their article "Spam!", which was published in the journal *Communications of the ACM*. Based on studies of 400 spam, they concluded that *money making opportunities* was the most common advertisement of spam, on second place was a category called *other products and services*

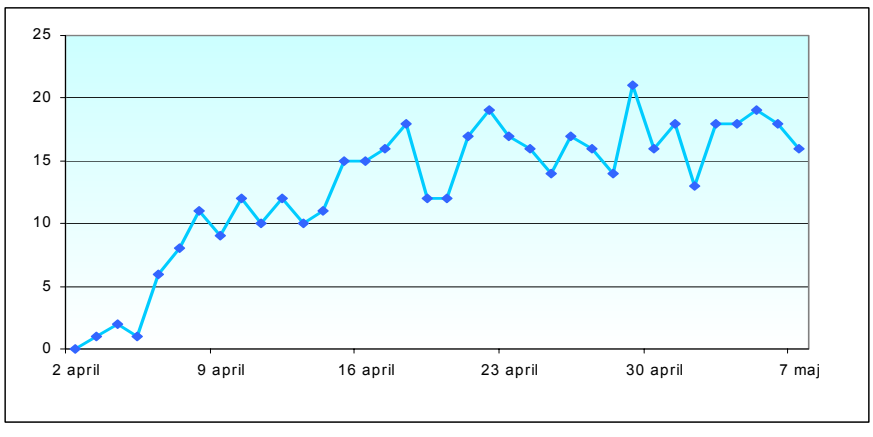

Figure 1 Number of spam day by day [2].

[7]. This category included phone services, vacation packages, nutritional supplements, weight loss products and on-line newsletters. In the experiment, the category, which is called free offers also match the result from Cranor and LaMacchia (this category called *other products and services*), because the advertisements in both categories are more or less of the same kind.

Cranor and LaMacchia also noted that only 36% of the messages contained instructions for how to be removed from the mailing list [7]. Perhaps most telling about the nature of these messages was the fact that fewer than 10 percent identified name, postal address, phone number, and e-mail address of the sender. As is described in Article 13 of the EU-Directive [6], this somewhat contradicts basic principles of privacy. Also, in the investigation, most of the spam offers the recipient the possibility to delete his och her e-mail address from the mailing list by using opt-out lists. Even though it is commonly known that the recipient should not reply to the sender of the spam or to sign any opt-out lists (then, the spammer will know that the e-mail address is active and possibly spam it even harder). However, in our investigation we could not find such a connection, that is; the unsubsription of spam did not result in getting new spam.

3 Final Remarks

In the above discussion we can conclude that there are three requirements for achieving sound e-commerce, of which the first one is crucially important. Consumers should be provided with a means to decide what, when, and for what purpose their personal information is being used. Without such a control of personal privacy, consumers will hardly provide companies with reliable personal data. If so, companies will most likely try to get consumer data elsewhere so that they can continue to access consumers with customised and/or personalised offers (e.g., via e-mail).

The experiments discussed in this paper have shown that consumers don't have the means to decide what, when, and for what purpose their information is used. There are some efforts on the regulatory area (e.g., the EU-Directive [6]), but even though it is carried into effect it has not yet had any real impact on the occurrence of spam. How-

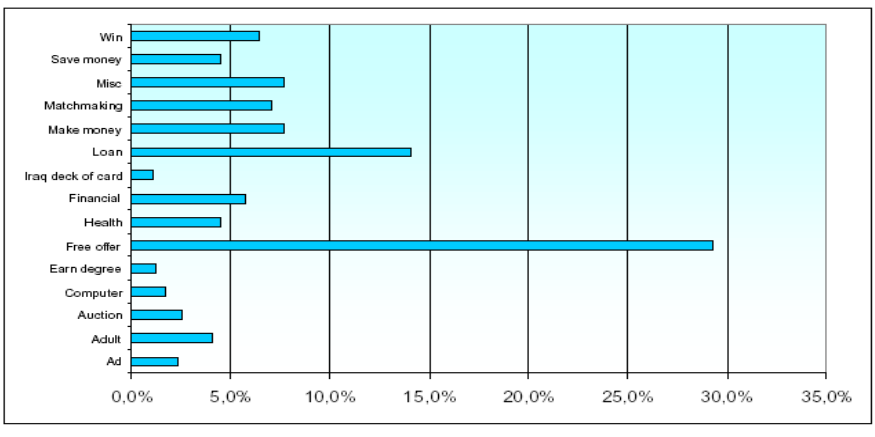

Figure 2 Spam categories [2].

ever, it might be a bit too early to assume that the Directive has no significance, given that it has only been in use since 2002. In addition, our investigations suggest that most serious companies actually do not send offers to consumers unless they have permission. Research indicates that most e-companies cease to send commercial messages when being unsubscribed to. Given this, the notion of the right to be let alone seems to be not so far away. On the other hand, statistics concerning the occurrence of spam say otherwise. Over the last couple of years the amount of spam has augmented, and one statistical institute [1] predicts that spam will increase in number over the next years.

In the experiment by Gunnarsson and Ekberg privacy policies for every visited site were collected for a later evaluation. From this point of view the information collection about users was mentioned in the fine print of the privacy policy. The spam list generated was selective, an American address was included but not a Swedish one. The spam sent have general contents, not connected to the music site.

The spam community ranges from close to ordinary customer relations, to "snake oil" messages. There is a fine line between what users or customers regard as useful information and what is intrusion to personal privacy. One thought is that the more personalised the offers are, the more likely users are to regard them as privacy invaders. If so, what happens when offers arrive to end users in such an extent that they hardly are able to distinguish personal messages, and possibly serious offers from all the offers. So, there should be a great risk for the success of e-commerce if the volume of unsolicited e-mails continue to grow without discrimination.

Tragedy of the commons occurs when there is a conflict of interest between a single actor and the whole community [3]. The situation arises if the benefit of the single actor exceeds the benefit of belonging to the community. Albeit the investigations have shown that most of the web sites behaved well, a small fraction generated a large amount of spam. This is the tragedy of the commons.

References

[1] Ferris Research; http://www.ferris.com/, 2003-06-16

[2] Gunnarsson, A. and Ekberg, S. Invasion of Privacy, Master Thesis BTH, 2003

[3] Hardin, G., "The tragedy of the commons", Science vol. 162 pp. 1243-1248, 1968

[4] Sterne, J., and Priore A., E-Mail Marketing - Using E-Mail to Reach Your Target Audience and Build Customer Relationships, John Wiley & Sons, Inc., USA, 2000

[5] Fischer-Hübner, S., IT-Security and Privacy - Design and Use of Privacy-Enhancing Security Mechanisms, Springer-Verlag; Lecture Notes in Computer Science; vol. 1958; Germany, 2001

[6] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications); http://www.cdt.org/privacy/guide/protect/telecom-priv02.pdf, 2002

[7] Cranor, L.F., and LaMAcchia, B.A., "Spam!", 1998, Communications of the ACM, Vol. 41, No. 8 (Aug. 1998), Pages 74-83, Definitive version: http://lorrie.cranor.org/pubs/spam/spam.html, 1998