

The Economy of Spam

Michael Ilger Jürgen Strauß Wilfried Gansterer
Christian Proschinger

September 12, 2006

Technical Report FA384018-6
Institute of Distributed and Multimedia Systems
University of Vienna

Abstract

Although many different tools are already available which lessen the burden of spam, it still remains a severe problem. A very promising solution for preventing *outgoing* spam starts at a different point. Spammers should be driven out of business by creating cost high enough to make their work unprofitable. In this report we try to show under which circumstances this goal can be reached and we present a tool for analyzing spammers' business models.

1 Motivation

For the last couple of years every e-mail user had to face the same problem: An ever increasing number of spam messages. As it is a widely accepted fact that spam costs a lot of money, and additionally is very annoying, many different approaches have been proposed to decrease the amount of spam, all of them with certain advantages and disadvantages. The most common approach is using a filter for detecting spam messages *after* they have been delivered. Usually this approach is based on certain rules, a Bayes filter, or a combination of both [12]. Provided that all included elements receive the proper tuning and training such filters can achieve a high detection rate. They have the advantage of being fairly simple, but the great disadvantage of taking effect only very late, after the message is already accepted by the receiver. Other approaches, like greylisting [8], have the advantage of taking effect a little earlier while providing a similar filtering performance but have the disadvantage of introducing delays in the mail delivery process.

Instead of filtering messages it is also possible to try to prevent spam messages from being sent by using new protocols or including authentication mechanisms in existing protocols [9]. While such a solution promises good results,

it has the extreme disadvantage of requiring a complete replacement of all e-mail software globally, which is unrealistic. Finally, a number of laws has been passed to prosecute spammers [10] [13], and newspapers have featured headlines about spammers being caught and brought to trial. Still, this did not solve the problem.

One important aspect has not received the amount of attention which it deserves. While only a small number of spam messages is sent out “just for fun”, the vast majority of spam has one simple purpose: to return profit. Currently, with messaging cost being close to zero, it is possible to create profit with a very low response rate (see Section 4.3). The goal in this report is to take a closer look at the cost a spammer is facing when sending out a certain amount of messages. We will also try to relate this to the cost a regular user is facing when using e-mail in order to find economical barriers which impede spammers while being unnoticed by a regular user [5]. With the knowledge gained here it is possible to create new tools for outgoing spam prevention which can help internet service providers (ISPs) to reduce the load on their servers as well as to secure their reputation by preventing spamming originating from their network.

2 Analyzing Spammers’ Business Model

Obviously spammers want to make profit. The only way to stay in business is to create income which is higher than the total expenses. This leads to two simple options for stopping spam: Either restricting the potential income from spamming (but this is hard to achieve), or raising the cost for sending out spam high enough to make such business unprofitable. The different business models motivating the spam phenomenon have already been summarized in [3] [4]. In the following sections we will focus on an overview of the bills a spammer has to pay at the end of the day as well as on different approaches for raising his cost.

2.1 Cost-based Anti-spam Approaches

While money is the first thing that comes to mind when talking about “cost”, cost-based anti-spam approaches do not necessarily involve money directly. Even though creating (paid) stamps for e-mail messages is one of the existing approaches [1] [11] (with the money preferably donated to charity), it is not the only way of payment. In [5] we adapted a token-bucket strategy, which is traditionally used for traffic shaping, to fit the needs of e-mail traffic regulation. It is based on creating its own “currency” (*tokens*) for e-mail messages and includes a starting seed as well as a certain refill rate. Each message sent consumes a certain number of tokens. This implementation allows to impede heavy-duty users while remaining unnoticed by common users.

Similarly, other approaches require users to solve certain puzzles to be able to send a message. These puzzles may depend on human interaction or require

the computer to do certain computations [6]. In both situations human users can easily meet the requirements whereas mass-mailers cannot.

2.2 Cost Factors for Spammers

Before different cost and profit models can be analyzed, the most important cost and revenue factors must be identified. The cost factors for a spammer can be grouped in four categories—hardware cost H , software cost S , operating cost O , and labor cost L . Figure 1 provides an overview of the most important cost factors. Some cost, such as hardware cost, are easy to measure, but others can only be estimated (software installation duration, time to compose a spam e-mail). The basic cost model for a spammer is thus defined as:

$$\text{total cost } c = H + S + O + L.$$

Assuming a single spammer uses his own home equipment, hardware cost H can be defined as the sum of the cost C for a computer, M for a monitor, and P for peripheral devices:

$$H = C + M + P.$$

Software cost S can be divided into cost for basic software which every computer needs (like cost OS for the operating system) and cost for software for special spamming activities, like cost R for remailers, cost MAH for mail address harvesters or cost WH for web hosting:

$$S = OS + R + MAH + WH.$$

Operating cost O is a sum of internet service cost I , electricity cost E for running the system, address collection cost A (addresses can be bought or self-collected) and open proxy cost OP :

$$O = I + E + A + OP.$$

Labor cost L can be divided into cost IN for installation, cost MT for maintenance, cost MP for mail production, and cost AC for acquiring customers:

$$L = IN + MT + MP + AC.$$

All these cost factors must be measured or estimated for a certain period of time (for example, per day). The resulting daily cost is then divided by the number of messages which are sent out in this period of time in order to determine the per message cost.

2.3 Revenue Factors for Spammers

While the cost factors can be estimated relatively easily, revenue factors are very hard to define and determine, because spammers usually do not declare their business model in public. Nevertheless, it is clear that there are two main

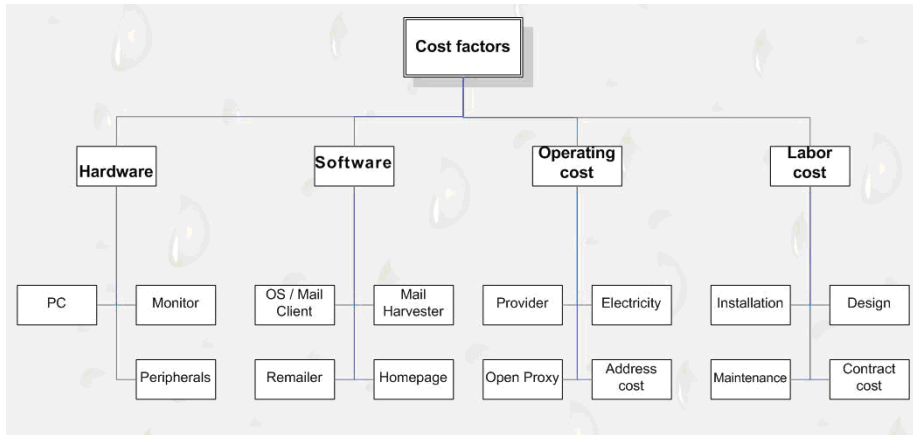


Figure 1: Cost factors for spammers

payment schemes—a brokerage system, where the marketer gets a fee per item sold and a pay-per-mail campaign system (the spammer receives money for an e-mail campaign to a certain number of customers). As it is not possible to get representative and reliable data about the former the investigation presented here focuses on the mail campaign system. As discussed in [4], a good estimate for the average revenue per message sent out is 0.00434 Euro.

3 The SpamSim Tool

In order to design and to calibrate an anti-spam method which is capable of interfering with these business models, we created the simulation tool *SpamSim*. This tool allows us to model the effects of various parameters on the profit achieved by spammers.

3.1 System Overview

Our tool implements the most important cost and profit factors and models their interdependencies. Based on this tool, we try to answer central questions, such as how many messages must be filtered out, or how many messages a spammer has to send to be profitable. By examining the break-even point for spammers in terms of cost and profit, we can evaluate the effectivity of anti-spam methods.

The SpamSim tool has been implemented as a Windows GUI application, programmed in Visual C# operating on the Microsoft .NET framework. The interface is currently only available in the German language. Future versions may also provide the logic of this program as a web-service with an English user interface. To provide a flexible solution the scenarios are defined in MathML—each scenario is defined by the elements of the cost factors. MathML (Mathematical Markup Language) is an application of XML which includes additional elements

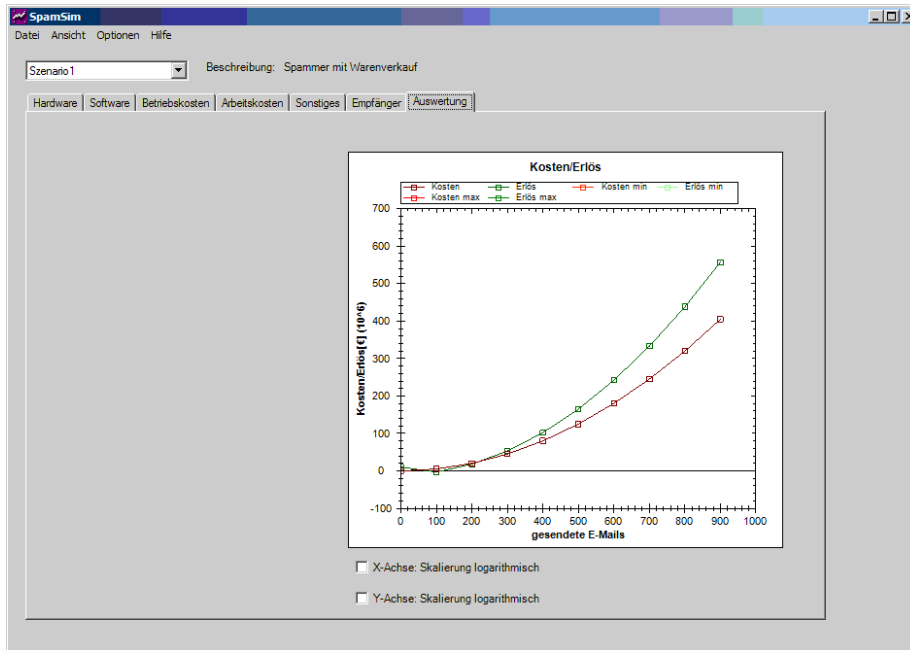


Figure 2: Graphical output of SpamSim

for calculations. The SpamSim tool has a standard and an advanced mode—the advanced mode allows fine tuning of all parameters, whereas the standard mode uses default values for many parameters. The results of the simulation can be displayed as a line diagram or exported as an XML or CSV file.

3.2 Configuration Details

SpamSim consists of an executable file, a library which is used to create graphics, and a configuration file. The config file provides the path to further input files. One of this input files determines the factors and default values used for the computation, while three more for each scenario provide the formulae used to compute the values shown on the x and y axis of the diagrams.

The program itself features four different views. These views include a preferences window to choose the right configuration and an editor which allows for easy access to the scenario as an xml file. The cost can be modeled in the main window. Finally, a scenario window as shown in Figure 2 provides the graphical solution.

3.3 Tool Usage

Using SpamSim is relatively simple. It does not require an installation, but instead only an entry for the installation path in the configuration file. The

currently easiest and fastest way to create an xml file which is used as input for the program is to modify an existing file with a regular text editor. Optionally, parameter values may also be edited within the application. One file contains all the scenarios with their parameters and default values. Its file path is part of the configuration file. Optionally it is possible to save and load the values of a certain scenario. The calculations for each scenario are defined in a total of three different files. Each filename begins with the scenario name and is followed by a certain usage convention. One is used to define the stepping along the x-axis. The other two files include the cost and revenue formulas for the scenarios defined earlier.

In the following section, three different scenarios are discussed in more detail. All of them were implemented with SpamSim, and the figures shown for illustration were created with the tool.

4 Selected Case Studies

In this section we describe three different scenarios of spamming activity and possible solutions to destroy spammers' business models. It is clear that the profit of a spammer depends on the number of messages sent out. First we analyze the influence of a token bucket algorithm as described in [5] which dynamically limits outgoing spam traffic (from ISP side), second we investigate the impact of filter performance and third we explore the leverage of the response rate on spammers' business.

4.1 Scenario 1

This scenario provides a simple example of a spammer who uses his own home PC and a leased line for spamming. The corresponding expense factors are listed in Table 1. We assume that hardware, software, operating cost and working cost are monthly fixed cost. Only open proxy cost, which are paid per e-mail sent out, can be denoted as running cost.

If the upload line has a bandwidth of approximately 256 KBit/sec, if we assume an average size of a spam message around 4 Kbyte (this average size is based on the analysis of approximately 1 000 000 spam messages [2]), and if we ignore all overhead data, a spammer is able to send out around 691 200 messages per day (or 20 736 000 messages per month). Figure 3 shows the cost and the revenue in this scenario. The cost curve is taken from the above data, and the revenue curve is based on a revenue of 0.00434 Euro per e-mail [4]. Figure 3 clearly demonstrates that the spammer can make a huge profit if there is no restriction.

Restricting the outgoing traffic to 10 000 messages per day leads to the diagram shown in Figure 4. It can be clearly seen that the cost and revenue curves intersect close to 250 000 messages sent out per month. Below the spammer does not make any profit.

Type of cost	Total cost [Euro]	Life time [month]	Monthly cost [Euro]
<i>Fixed cost</i>			
Hardware			
Average computer	1000.-	36	27.78
Average Monitor	300.-	36	8.34
Peripherals	200.-	36	5.55
Software			
Operating system (open source)	1.-	36	0.027
Harvester	50.-	36	1.38
ReMailer	100.-	36	2.77
Homepage	20.-	1	20.00
Operating cost			
ISP cost	49.-	1	49.00
Electricity	35.-	1	35.00
Addresses	300.-	12	25.00
Working cost			
Installation	30.-	12	2.50
Maintenance	300.-	1	300.00
Mail-design	300.-	1	300.00
Customer recruitment	300.-	1	300.00
sum of fixed cost:			1077.36
<i>Variable cost</i>			
Open proxy	0.000125	per mail	

Table 1: Cost factors for a single spammer (scenario 1)

e-mail sent [$\times 10^3$]	cost [Euro]	revenue [Euro]
0	1077.36	0.00
25	1080.48	108.50
50	1083.61	217.00
75	1086.73	325.50
100	1089.86	434.00
125	1092.98	542.50
150	1096.11	651.00
175	1099.23	759.50
200	1102.36	868.00
225	1105.48	976.50
250	1108.61	1085.00

Table 2: Cost factors single spammer scenario 1 limit=8400

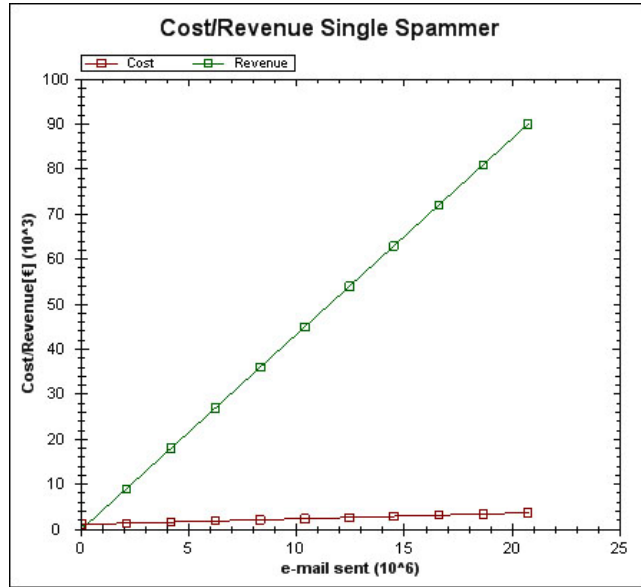


Figure 3: Cost (printed in red) and revenue (printed in green) spammer (scenario 1, no limit)

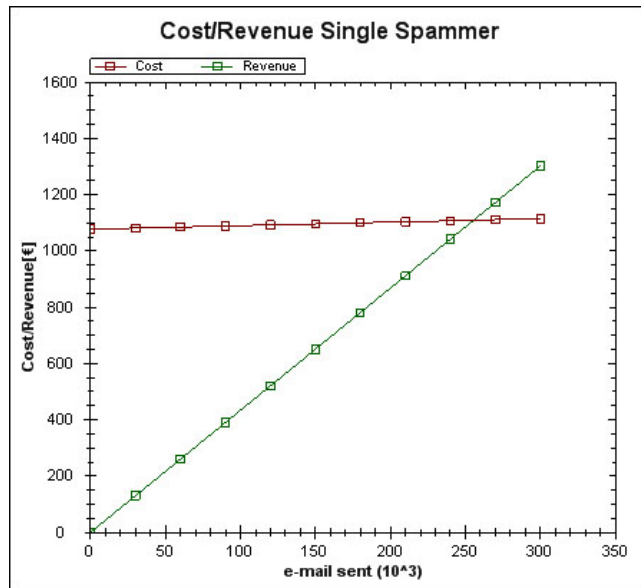


Figure 4: Cost (printed in red) and revenue (printed in green) spammer (scenario 1, limited to 10 000 messages per day)

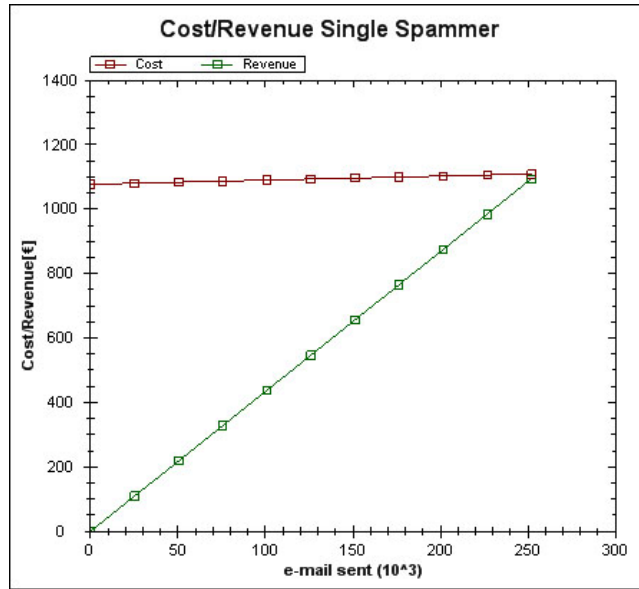


Figure 5: Cost (printed in red) and revenue (printed in green) spammer (scenario 1, limited to 8 400 messages per day)

Table 2 and Figure 5 show the cost and revenue data for a limit of 8400 messages. In this case, the revenue always stays below the cost. Consequently, in the given scenario (single spammer) we find that a restriction of outgoing traffic to less than 8 400 messages per day will destroy spammers’ business model.

4.2 Scenario 2

This scenario describes a single spammer, who uses a leased line to connect to a leased server called LegalMail.¹ The spammer connects remotely to this server. The server then sends out user defined messages. Table 3 shows the cost factors for this scenario. Unlike scenario 1 there are no fixed cost—all the monthly cost are divided through the amount of e-mail sent and afterwards compared to the revenue.

LegalMail guarantees that under certain conditions (size of the messages, not too many invalid addresses) a minimum of 30 000 000 messages is sent out per month. The total monthly cost are 2 018.42 Euro. Thus, one message costs 0.000067 Euro, but brings 0.00434 Euro of revenue. As Figure 6 shows that the revenue can be huge if no countermeasures are taken.

In this scenario we show how important the *spam filter performance* is for harming spammers’ business. The better the filter performance the smaller is the profit. Quantitatively speaking, with a filter performance of 90% spammers

¹<http://www.americaint.com/bulk-email-software/legalmail/legalmail.html>

Type of cost	Total cost [Euro]	Life time [month]	Monthly cost [Euro]
<i>Fixed cost</i>			
Hardware			
Average computer	1000.-	36	27.78
Average Monitor	300.-	36	8.34
Peripherals	200.-	36	5.55
Software			
Operating system (open source)	1.-	36	0.027
Homepage	200.-	36	5.55
Operating cost			
ISP cost	35.-	1	35.00
Electricity	17.-	1	17.00
Addresses	200.-	12	16.67
LegalMail	1000.-	1	1000.00
Working cost			
Installation	30.-	12	2.50
Maintenance	300.-	1	300.00
Mail-design	300.-	1	300.00
Customer recruitment	300.-	1	300.00
sum of fixed cost:			2018.42

Table 3: Cost factors single spammer (scenario 2)

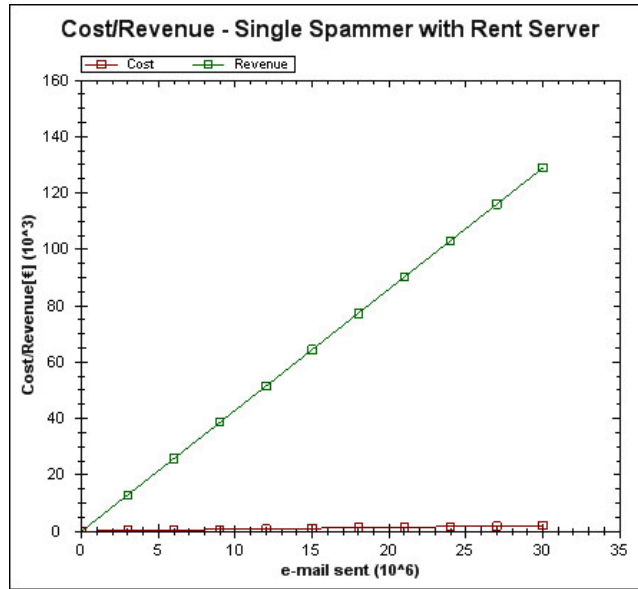


Figure 6: Cost (printed in red) and revenue (printed in green) spammer (scenario 2, no spam filter)

are still making profit (see Figure 7). Only if the filter achieves a detection rate of 97,45% or more, spamming becomes unprofitable (see Figure 8 and Table 4).

4.3 Scenario 3

This scenario describes the same spammer as in scenario 1, but this spammer does not get his revenue from sending out e-mail. Instead, he advertizes by e-mail and is paid per item sold. In this example we describe the influence of the *response rate* on the spammers' business model. Let us assume that the spammer gets a provision of 8 Euro per item sold and that the response rate is 0.036% (the spammer sells one item per 2777 messages sent out [3]). Then the resulting graph looks as shown in Figure 9. The same scenario with a lower response rate of 0.001% is shown in Figure 10.

With a response rate of 0.00001% (cf. [7]) the spammer could never generate any profit in this scenario. Table 5 shows that the break-even point is close to a response rate of 0.002% (which corresponds to one out of 50 000 messages leading to a successful business transaction). A discussion of estimated response rates is also included in [4].

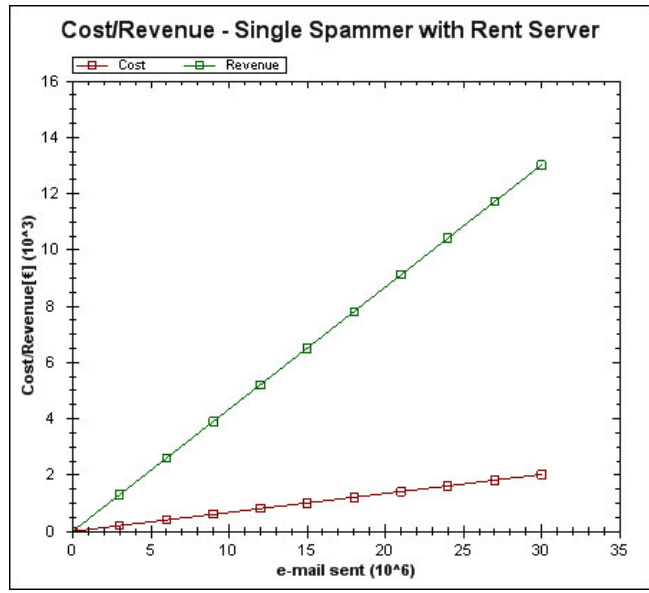


Figure 7: Cost (printed in red) and revenue (printed in green) spammer (scenario 2, spam filter with detection rate 90%)

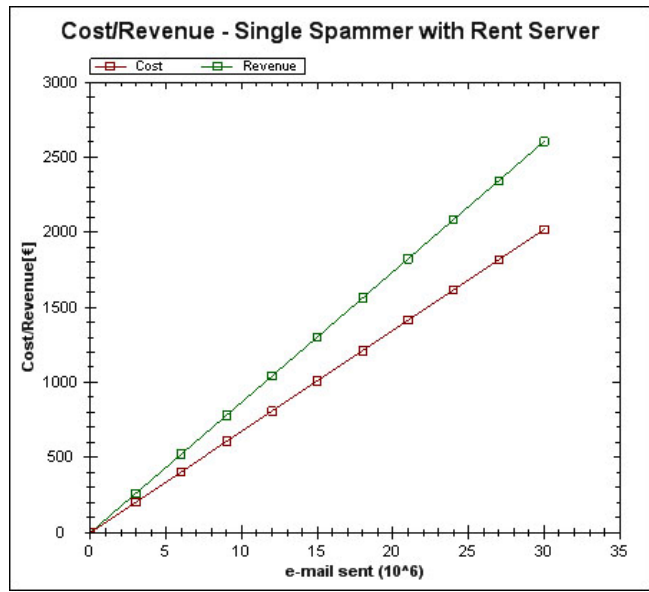


Figure 8: Cost (printed in red) and revenue (printed in green) spammer (scenario 2, spam filter with detection rate 98%)

e-mail sent [$\times 10^9$]	cost [Euro]	revenue [Euro]
3	201.84	201.81
6	403.86	403.62
9	605.53	605.43
18	1211.05	1210.86
27	1816.58	1816.29
30	2018.42	2018.10

Table 4: Cost factors single spammer (scenario 2, spam filter with detection rate 97.45%)

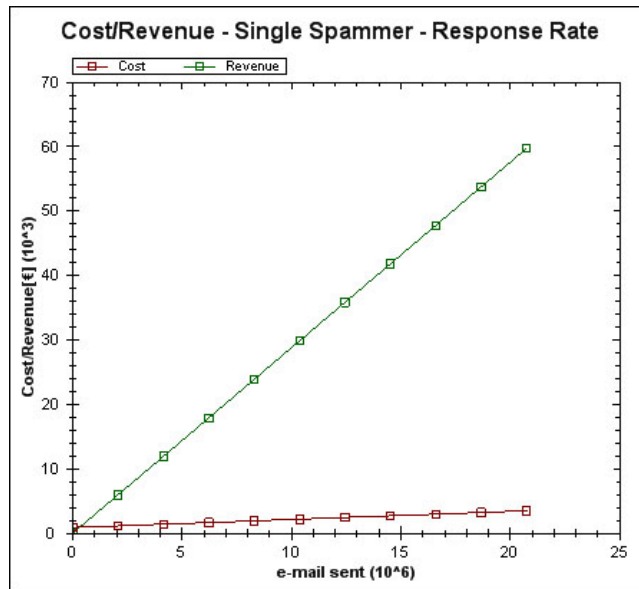


Figure 9: Cost (printed in red) and revenue (printed in green) spammer (scenario 3, response rate 0.0036%)

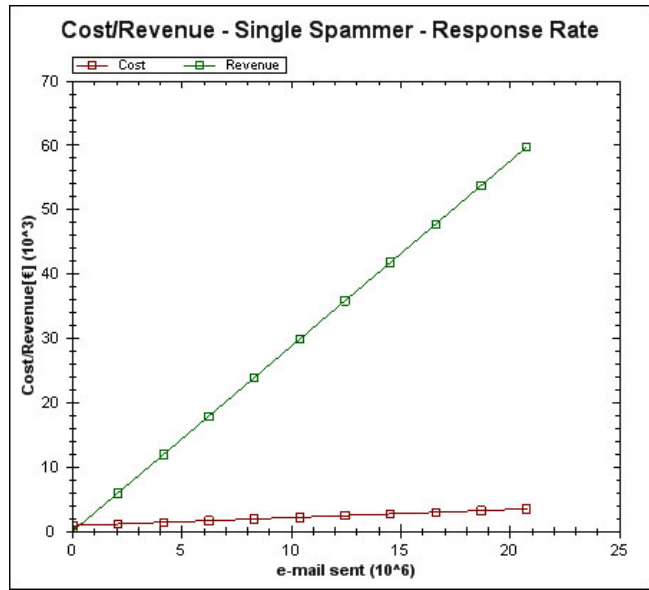


Figure 10: Cost (printed in red) and revenue (printed in green) spammer (scenario 3, response rate 0.001%)

e-mail sent [$\times 10^9$]	cost [Euro]	revenue [Euro]
0.00	883.19	0.00
2.07	1142.39	331.78
4.14	1401.59	663.55
6.22	1660.79	995.33
8.29	1919.99	1327.10
10.36	2179.19	1658.88
12.44	2438.39	1990.66
14.54	2697.59	2322.43
16.59	2956.79	2654.21
18.66	3215.99	2985.98
20.73	3475.19	3317.76

Table 5: Cost factors single spammer (scenario 3, response rate 0.002%)

5 Conclusions and Future Work

Detailed information on the relevant cost and revenue factors in the spammers' business model is very important when trying to stop spam, especially as it allows to predict which anti-spam strategies can be expected to be effective.

In this report, it has been shown how to utilize this information for evaluating anti-spam methods and strategic decisions. In particular, the SpamSim tool has been presented which allows for quantitative modelling the business model underlying spamming. Using this tool, the economic aspects of three different spamming scenarios and anti-spam strategies (restricting outgoing traffic, filtering, reducing response rates) have been analyzed.

Based on this work, we plan to analyze the cost caused by spam from an overall economic point of view which also includes the cost created by spam-prevention. Factoring in detailed information on various cost factors also allows for a new approach to many pre-send techniques. This is expected to lead to a fine tuning of these approaches which boosts their performance. This in turn will make it much harder for spammers to circumvent these anti-spam methods and at the same time makes them fully transparent to regular users.

Acknowledgments. We would like to express our gratitude to Internet Privatstiftung Austria, mobilkom austria, UPC Telekabel, and Internet Service Providers Austria for supporting this research.

References

- [1] CAPEK, P., LEIBA, B., WEGMAN, M., AND FAHLMANN, S. Charity begins at ...your mail program, 2004. [http://domino.research.ibm.com/comm/research_projects.nsf/pages/spam.papers.html/\\$FILE/charity-seals.pdf](http://domino.research.ibm.com/comm/research_projects.nsf/pages/spam.papers.html/$FILE/charity-seals.pdf).
- [2] DABERGER, M. Quantitative Untersuchungen der Eigenschaften von Unsolicited Bulk E-Mail und Unsolicited Commercial E-Mail. Master's thesis, Institute of Distributed and Multimedia Systems, Faculty of Computer Science, University of Vienna, June 2006.
- [3] GANSTERER, W., ILGER, M., LECHNER, P., NEUMAYER, R., AND STRAUSS, J. Anti-spam methods—state of the art. Technical Report FA384018-1, Institute of Distributed and Multimedia Systems, Faculty of Computer Science, University of Vienna, Mar. 2005.
- [4] GANSTERER, W., ILGER, M., LECHNER, P., NEUMAYER, R., AND STRAUSS, J. Phases 2 and 3 of project 'Spamabwehr': SMTP based concepts and cost-profit models. Technical Report FA384018-2, Institute of Distributed and Multimedia Systems, Faculty of Computer Science, University of Vienna, 2005.
- [5] GANSTERER, W. N., HLAVACS, H., ILGER, M., LECHNER, P., AND STRAUSS, J. Token buckets for outgoing spam prevention. In *Proceedings of*

the IASTED International Conference on Communication, Network, and Information Security (CNIS 2005) (Nov. 2005), M. Hamza, Ed., IASTED, ACTA Press.

- [6] GOODMAN, J. T., AND ROUNTHWAITE, R. Stopping outgoing spam. In *ACM Conference on Electronic Commerce* (2004), pp. 30–39.
- [7] GRAHAM, P. A plan for spam, 2002. <http://www.paulgraham.com/stopspam.html>.
- [8] HARRIS, E. The next step in the spam control war: Greylisting. Tech. rep., PureMagic Software, 2003. <http://projects.puremagic.com/greylisting/whitepaper.html>.
- [9] LECHNER, P. Das Simple Mail Transfer Protokoll und die Spamproblematik. Master's thesis, Institute of Distributed and Multimedia Systems, Faculty of Computer Science, University of Vienna, June 2005.
- [10] LEE, Y. The can-spam act: a silver bullet solution? *Commun. ACM* 48, 6 (2005), 131–132.
- [11] LODER, T., VAN ALSTYNE, M. W., AND WASH, R. An economic answer to unsolicited communication. In *ACM Conference on Electronic Commerce* (2004), pp. 40–50.
- [12] SCHRYEN, G. Effektivitaet von Loesungsansetzen zur Bekaempfung von Spam. *Wirtschaftsinformatik* 46, 4 (2004), 8.
- [13] SIPIOR, J. C., WARD, B. T., AND BONNER, P. G. Should spam be on the menu? *Commun. ACM* 47, 6 (2004), 59–63.