

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Thesis

USING MARKETS AND SPAM TO COMBAT MALWARE

by

SARAH LIEBERMAN ZATKO

B.S., Massachusetts Institute of Technology, 2005

Submitted in partial fulfillment of the
requirements for the degree of
Master of Arts
2009

Approved by

First Reader _____
Stephen Homer, PhD
Professor of Computer Science

Second Reader _____
Azer Bestavros, PhD
Professor of Computer Science

Acknowledgements

Marshall van Alstyne is a co-author of this paper. I performed this work as his Research Assistant.

We are grateful to Mark Reynolds for his assistance in clarifying the ideas presented here. We also appreciate the work that Theodore Loder and Rick Walsh have done on the attention bond model. Thanks to Virgilio Almeida as well, for his useful feedback on this paper.

A preliminary version of this work was presented at the workshop on Interdisciplinary Studies in Information Security at Monte Verita in Ascona, Ticino during July 2008. The completed work was presented at WISE 2008: Twentieth Workshop on Information Systems and Economics in Paris, France, in December 2008. “Markets Can Cure Spam Zombies Too” won best student talk at MIT Spam 2009.

USING MARKETS AND SPAM TO COMBAT MALWARE

SARAH LIEBERMAN ZATKO ¹

Abstract

We propose an economic mechanism to reduce the incidence of malware that delivers spam. Earlier research proposed attention markets as a solution for unwanted messages, and showed they could provide more net benefit than alternatives such as filtering and taxes. Because it uses a currency system, Attention Bonds faces a challenge. Zombies, botnets, and various forms of malware might steal valuable currency instead of stealing unused CPU cycles. We resolve this problem by taking advantage of the fact that the spam-bot problem has been reduced to financial fraud. As such, the large body of existing work in that realm can be brought to bear. By drawing an analogy between sending and spending, we show how a market mechanism can detect and prevent spam malware. We prove that by using a currency (i) each instance of spam increases the probability of detecting infections, and (ii) the value of eradicating infections can justify insuring users against fraud. This approach attacks spam at the source, a virtue missing from filters that attack spam at the destination.

Additionally, the exchange of currency provides signals of interest that can improve the targeting of ads. ISPs benefit from data management services and consumers benefit from the higher average value of messages they receive. We explore these and other secondary effects of attention markets, and find them to offer, on the whole, attractive economic benefits for all – including consumers, advertisers, and the ISPs.²

¹Co-Authored with Marshall van Alstyne

²This material is based upon work supported by the National Science Foundation under Grant No. 0114368. Funding was also provided by a Boston University Dean's research fellowship.

Contents

1	Introduction	1
2	Previous Work	2
3	Botnet Background and Previous Work	5
4	The Model	8
4.1	Perfect Filter	9
4.2	Attention Bond	11
5	Information Security Effects	11
5.1	Detection	12
5.2	Prevention	15
6	Advertisers	19
7	Discussion and Possible Extensions	23
8	Conclusions	25
	References	28

List of Abbreviations

ABM	Attention Bond Mechanism
CAPTCHA	Completeley Automatic Public Turing Test to Tell Computers and Humans Apart
C&C	Command and Control
CPU	Central Processing Unit
CR	Challenge Response
DDoS	Distributed Denial of Service
HTTP	Hypertext Transfer Protocol
IRC	Internet Relay Chat
ISP	Internet Service Provider
P2P	Peer to Peer

1 Introduction

The Attention Bond Mechanism (ABM) is a proposed solution for spam. When a message is sent, the sender posts a minimal amount of money as his pledge, or “bond” that the message is not spam. If the recipient of the message agrees that the message is of interest, he ignores the bond which simply returns to the sender. On the other hand, if the message is unwanted the recipient claims it, thus punishing the sender for lying about message content. This system is primarily meant to cover first-contact communication. Once the initial contact has been made a sender can be whitelisted, and bonds avoided in the future.

This system was proposed in a previous paper (Loder *et al.*, 2006) and compared favorably to two of the dominant proposals for reducing spam. A critique of the ABM, however, argues that malware would make the system economically infeasible (Timothy, 2005; Lim, 2008). We address that issue here. We show that not only does malware not pose a significant problem for the ABM, but with minor adjustments ISPs can leverage attention bonds to detect and prevent malware infections. The end result is that aspects of ABM, which appear initially as weak links in the system, in the end turn out to offer very exciting results.

As a primary effect, the ABM has the potential to reduce the spam messages end users receive. Its secondary effects, however, could cut down on the volume of messages ISPs must filter, create new profitable markets for the ISP, cut costs to advertisers, and decrease the level of botnet infection among end users.

In section 2 that follows, we lay out background in anti-spam research. Section 3 then sets the scene regarding the problem of botnets and malware. In Section 4 we describe the model we use to prove the feasibility of the ABM. Information security is the topic of Section 5 where claims of feasibility and profitability are proven in Subsections 5.1 and 5.2. Ancillary effects on advertisers are discussed in Section 6, while broader implications

and possible future work are covered in Sections 7 and 8 respectively.

2 Previous Work

There have been many different types of solutions proposed for the problem of spam³. The ones which are most in use today are technological solutions, such as identity-based or content-based filtering. Blacklists and whitelists involve rejecting or allowing access respectively, based on the sender's membership in a list maintained on the receiver's end. Blacklists are problematic due to the cheapness of new identities. It is easy for spammers to create new identities each time an old one becomes blocked. This is symptomatic of open societies - ones where there is no charge per access - and results in a need for newcomers to earn a good reputation. (Friedman & Resnick, 2001) Whitelists are similar, except that the list which is maintained identifies those who are allowed, and all others are dropped. (Cranor & LaMacchia, 1998) These are vulnerable to the spoofing of identities, but can be useful if authentication is used to make identities "strong". (Tompkins & Handley, 2003) However, they do not create an easy mechanism for first-contact situations. The attention bond solution provides this first-contact mechanism, and then uses whitelists for subsequent contacts.

Content-based filtering tries to classify a message as spam based on the content of the message. This is intuitive, as it makes sense to try and figure out what a message is based on what it says, but in practice there are a number of problems with this approach. Firstly, different people define spam differently (Rainie & Fallows, 2004), and this requires everyone to agree on which messages are wanted and which are not. Much more worrying is the actual analysis needed. It is very easy for a sender to use metaphor, intentional

³Anti-spam papers include: Dwork & Naor (1993); Laurie & Clayton (2004); Fahlman (2002); Hermalin & Katz (2004); Krishnamurthy (2004); Pantel & Lin (1998); Sahami *et al.* (1998); Tompkins & Handley (2003); von Ahn *et al.* (2003).

misspelling, and various other tools to obfuscate a message's content. In order to guarantee that no false positives or false negatives result, the filter would have to essentially pass a Turing test. Anything short of this cannot deal with the near-infinite creativity of spammers trying to get a message through. Another nuance of this technology is that the same filters which are used to protect recipients can be used by senders to test messages and determine which will get delivered. (Graham-Cumming, 2004) Filtering of this sort suffers from an imbalance of information, because only the sender knows if the message is spam. ABM, on the other hand, relies on the sender's knowledge of the message content. This is why ABM does not require any of the advanced content analysis needed for filtering.

“Community filters” have a slight advantage in deciphering the meaning of messages because they harness the parsing ability of users. Some users categorize messages as being either spam or legitimate. This knowledge is then used to better filter similar messages for all the other users of the system. This method still suffers from false positives and false negatives, both due to mis-categorizations and to differing opinions regarding the definition of spam.

Regulatory efforts, such as the CAN-SPAM Act of 2004, have been also made to combat spam. CAN-SPAM required senders to include valid subject lines, return addresses, and labels for any adult content. The legislation had very little effect on the overall problem, due largely to enforceability and jurisdiction issues. (Rainie & Fallows, 2004) Email easily passes across jurisdictional boundaries, so rules would have to be agreed upon by multiple governments and agencies. The larger problem is that a great deal of spam is not sent from the spammers' machines themselves, so accountability is very hard to achieve. Without accountability for messages, there is no way to prosecute those who break anti-spam laws. The difficulty of finding spammers is so great that one legal scholar even suggested that the federal government place bounties on criminal spammers. (Bazeley, 2003) The problem of botnets will be discussed in the next section, and this will shed further light on why

enforcement is such a hurdle to effective regulation.

These same problems also present a barrier to the effective deployment of tax mechanisms. If used properly, taxes can force senders to limit themselves and target messages.

There are market-based solutions other than attention bond which have been proposed. Stamps have had some success, in that they do cause greater selectivity on the part of senders. However, postage values were not seen by recipients to be a signal of value. (Kraut *et al.*, 2003) This would be a requirement for a deployed system, so more work would still be required.

First contact interactions are also sometimes handled by Challenge Response (CR) mechanisms, such as CAPTCHAs. (von Ahn *et al.*, 2003) Other variants require unrecognised senders to answer some computational challenge or show a proof of work. (Dwork & Naor, 1993) Each of these CR methods has its drawbacks. CAPTCHAs can be inverted; the spammer offers free access to adult content in exchange for someone answering the challenge which has been forwarded to their site. The other CR methods simply require work to be done, without any transfer of value or side payments. They are also vulnerable to attacks by zombies, which will be discussed in the next section. (Laurie & Clayton, 2004)

Mechanisms wherein “interrupt rights” are sold in order to allocate attention are an intriguing option, as this approach places value on the commodity which is really in question. (Fahlman, 2002; Ayres & Nalebuff, 2003) Some work has been done in this area, including van Zandt (2004), which analyzes a system wherein boundedly rational receivers use attention pricing in a Vickrey auction. ABM expands on this concept, in that it allows recipients to place a monetary value of their choosing on their own time and attention.

3 Botnet Background and Previous Work

In order to understand the spam problem, one must also be somewhat knowledgeable about the issue of botnets. A bot, or zombie, is a machine which has been compromised in such a way that the attacker now has the ability to make use of that machine's computing power himself. Bots are responsible for an estimated 65% of today's spam messages. (Rainie & Fallows, 2004; McPherson & Labovitz, 2007) Bots are also used for other purposes, such as distributed denial of service attacks (DDoS) and for collecting personal information usable for identity theft.

In addition to being used for sending spam, bots also contribute to the failures of some of the anti-spam attempts from the previous section. Purely computational challenge-response or proof of work mechanisms can be thwarted by using the computational power of a botnet. Additionally, the already significant enforcement issues facing regulatory attempts are further hampered by the fact that offending messages are being sent from machines whose legal owners are completely unaware of their involvement in the matter. Botnets make tracking spam messages back to the true originators even more difficult.

Many bots will work in coordination as a botnet, directed by a command and control (C&C) mechanism. The specific protocols used to carry command and control traffic may vary, as well as the command structure within the botnet. Some are distributed, while others are centralized. Protocols used include IRC, HTTP, and P2P protocols.

Botnet infection is quite common, and thus there is a great deal of interest in developing new techniques for detecting and preventing such compromises (Jung, 2006). Recent studies show that botnets have become the biggest concern among ISPs, relegating DDoS to the second spot. As was just mentioned, these concerns are of course closely linked. Single botnet sizes have been reported in the tens and hundreds of thousands of infected machines. There are also indications that the volume of traffic resulting from bots is increasing at a

greater rate than the ISPs' capacity to handle that volume (McPherson & Labovitz, 2007).

Customer support and help desk calls create a significant cost to ISPs, and these costs are in part driven by the number of infected machines among the customer base. Poor end user security leads to infection, and thus to higher costs both from customer support and from fielding abuse notifications from other ISPs. (van Eeten & Bauer, 2008)

The picture this paints goes a long way to explaining why bots are important, but not why they are so hard to combat. Part of this problem is a lack of proper tools and resources. In interviews with security specialists across a varied spectrum of ISPs it was determined that the primary source of information used to find bots is the abuse reports that organizations receive from their peers. Such abuse reports will only get filed for the most egregious cases, so only a fraction of bots will be found this way. (van Eeten & Bauer, 2008) While it is important to pay attention to such reports, it is imperative that ISPs have a means of monitoring their own networks.

One of the factors which makes this problem difficult to resolve is an issue of moral hazard. The average computer user does not make use of the full computing power of their machine. Stolen CPU cycles will either occur only during idle time, or at most cause a minor slowdown. However, in most cases the user would not know what such a slowdown portends. Even if he did know that he had been compromised, the average user would probably not know what to do about it. The machine will be sending out spam and infecting other machines, but none of this will have any meaningful effect on the machine's user himself. This gives him little incentive to find or fix the problem.

The variations in specific methods used from one botnet to another also pose a difficulty. Many countermeasures only deal with some types of bots, and lack flexibility in the face of changes in how botnets operate. For example, Goebel & Holz (2007) uses n-gram analysis to identify bots based on signatures that had been detected in the nicknames for IRC bots. Binkley & Singh (2006) proposed another detection method specializing in centralized

IRC botnets, which instead searches for shared anomalies in TCP and IRC traffic among multiple machines.

Tools that are signature based are unfortunately easy to circumvent. As with content filtering for spam, these approaches will result in an arms race with the attacker. Every time signatures are updated, the botfarmers can change their behavior just enough so that they no longer match the signature.

The next generation of detection tools depend on finding correlations among multiple events and behaviors in order to detect infection. BotHunter uses intrusion detection systems with malware-specific sensors. Its detection is based on the infection lifecycle of a bot, and tracks the communication flows between a bot and its controller. (Gu *et al.*, 2007) BotSniffer identifies command and control traffic based on network anomalies. It primarily focuses on centralized networks. (Gu *et al.*, 2008b) BotMiner instead clusters similarities among communication traffic and malicious traffic to find hosts that are part of the same botnet. It identifies hosts that share the same anomalous patterns. (Gu *et al.*, 2008a)

These tools are a move in the right direction, but they still have their shortcomings. The amount of data which needs to be stored in order for the correlation engines to function does not scale well for large networks. The correlation engines can be fooled by introducing delays into the bot operations. In order to account for such delays, the tool in question would have to keep track of even more data than before, because a greater window of time would have to be covered. False positive rates also increase dramatically as the timeline correlation requirements are relaxed. These tools show promise, but issues such as these bar widespread deployment.

Any attempts to legislate against botnets run up against the same jurisdictional issues that make spam legislation so difficult. Malware is being used to support criminal endeavors, many of which are based overseas in poorer countries where a lower rate of job opportunities opens the door for criminal organizations. van Eeten & Bauer (2008)

ABM was criticized in the past for dealing with mail sent directly from the spammers' own machines, but not the more common messages sent from bots. We will discuss later why this factor, which initially appears to be a weakness, instead results in an even greater gain.

It is important to note that any organization that provides email accounts to its customers could implement the Attention Bond system. However, in this document we will refer to such an organization as an ISP, even though the entity may be of some other type.

4 The Model

It was shown in a previous paper that ABM can outperform perfect filters. (Loder *et al.*, 2006) That analysis did not account, however, for the existence of bots in the network. With bots it becomes possible for fraudulent charges to be made on the behalf of someone with a compromised machine. Here we will develop the model for comparing a perfect filter to the Attention Bond Model with the new security concerns included. We will focus on the economic implications of a policy wherein the ISP provides customers with protection against fraudulent charges. An initial proof of feasibility will expand to show that the detection of fraud can then be leveraged to detect botnet infection and even prevent it. First, though, the basic model must be developed.

There are certain variables that must be defined for the rest of this section. The net value of a message to the sender and receiver are s and r respectively. It should be noticed that while both are real numbers, s is non-negative. Any messages with negative s values do not get sent. The total number of customers being served by the ISP is N , and the rate of botnet infection among those customers is I , making for a total $N * I$ infected machines. A normal machine sends out m_n messages, while a machine infected by a spam bot sends out m_i . It can be assumed that $m_i \gg m_n$. Since an infected machine still

sends out the legitimate owner's mail as well, the number of spam messages is $m_i - m_n$. Variables which are specific to one case or the other will be defined as needed, but all of these values are relevant to both the filtering and bonding case. First to be discussed is the incumbent anti-spam mechanism, filtering.

Some assumptions are made in the following sections which should be explicitly stated for clarity's sake. First, in order to simplify the model we are assuming that the ISP's customers represent a closed system. That is, all messages go from one customer of the ISP to another customer of the same ISP. Several email providers could, for these purposes, be viewed as a single 'ISP', so this assumption does not place any undue constraints on the analysis.

We are also making an assumption that there is a one-to-one mapping between customers and machines. There is also an assumption made that all bots send spam. This is not strictly true, but the bots which are tasked with other purposes have no significant effect on the factors described here, and so are not of concern to us. Therefore, we will refer to bots and spam-bots interchangeably. None of these assumptions is strictly in line with reality, but making these assumptions should have no major effect on the results represented below. Finally, we define "attention bond fraud" here to be any incident in which a bond is posted by an account without the account owner's knowledge or permission.

4.1 Perfect Filter

The current standard for spam prevention is the filter. There are various different kinds of filters, and there is a constant competition between the people who send spam and those who try to filter it out. For the purposes of this comparison, we will disregard any problems that filters have with false positives and false negatives by using a "perfect filter" as the comparison control. A perfect filter would be one which blocks all messages that are

unwanted by the receiver, and which allows all others.

Filtering technology needs to be updated all the time, and new countermeasures are always being developed. Past studies have indicated that it can take as little as two hours from new filtering techniques appearing to spammers updating their systems accordingly. (Libbey, 2004) This is part of an ongoing arms race between security researchers and spammers, and there is no indication that a “silver bullet” solution will be developed for filtering any time soon. It is important to note that, although we are not representing it in the model here, filtering includes the ongoing cost of this cycle of development and deployment.

We are concerned here not with the value to any single entity, but rather the overall welfare created in the system, for both senders and receivers. W_f will be the value of that welfare in the filtering case. For the purposes of this analysis, we are assuming a closed system. That is, all senders and receivers are part of the ISP’s customer base.

The one value that appears in the equation below which has not yet been discussed is c_f . This is the cost of processing a spam message in a filtering system. Filtering mechanisms rely on analysis of the content of messages, so this cost will be non-trivial. Many companies are now applying filters and deep packet inspection to all of the messages they relay. (Detica, 2006; van Eeten & Bauer, 2008) The volume of messages which must be processed should result in a large processing load for such an endeavor, so any way to save these resources should be welcome.

$$W_f = (s + r - c_f)N \cdot m_n - N \cdot I(m_i - m_n)c_f \quad (1)$$

The first term is the net welfare per normal message times the expected total number of normal messages. The second is the expected number of spam messages times the cost per spam message. Remember that the infected machines are sending normal messages also,

so the number of spam messages from a single infected machine is $m_i - m_n$.

4.2 Attention Bond

Here, we view the welfare W_b which is achieved in the case where attention bonds are used. The cost c_b of processing a spam message in a bonding system should be far less than c_f , because the content of the message does not need to be processed at all. All the provider has to do is check whether a bond was posted, and only deliver the message if it was. The other variable unique to this equation is b , the average bond value set by receivers.

$$W_b = (s + r - c_b)N \cdot m_n - N \cdot I(m_i - m_n)(b + c_b) \quad (2)$$

The first term is the net welfare per normal message times the expected total number of normal messages. Note that here the net welfare also takes into account the additional cost of sending a bonded message. The cost per spam message has also changed, both to reflect the bond which is seized and the fact that the cost of processing a spam message is different in a bonding system than it is in a filtering system. It could be argued that the bond is transferring from one member of the system to another, and thus has no effect on the overall welfare. The scenario which we discuss in the upcoming section however deals with a case where the ISP subsidises fraudulent charges, so the bond transfer is a loss from the ISP, and thus cannot be ignored.

5 Information Security Effects

In the Attention Bond Model, spammers would most likely post a bond from one of the accounts on a compromised computer, and thus avoid losing their own money. This pro-

duces the potential for a problem very similar to the one of credit card fraud. Many of the methods used for dealing with credit fraud can be applied to this new kind of fraud as well. As is the case for credit cards, the ISP may choose to provide coverage to its customers in the case of fraudulent charges.

It was mentioned earlier that botnet detection is a problem due to multiple factors. The first obstacle to fighting bots is finding them. With traditional email there is no simple way to reliably track a message to its originating machine. This problem is resolved, because the exchange of a bond from one known and authenticated account to another creates an audit trail. It is easy now, when a spam message is received, to follow the seized bond back to the account that the message originated from. This also provides a means of identifying bots which is dependent only on the ISP's own resources, and thus provides them with a privacy-preserving means of monitoring their own networks, rather than relying on abuse reports from outside sources. (van Eeten & Bauer, 2008)

The other problem was the issue of moral hazard. Attention bonds take care of this as well. As it stood before, spam caused difficulties for others, but not for the spam-bot's owner himself. Now, each of those messages hurts him too, in the form of a seized bond. This change in incentives serves to eliminate the moral hazard.

5.1 Detection

The techniques which are used to detect credit card fraud and other cases of anomalous spending can be applied to attention bonds, producing some very favorable results. This detection can be based on changes in the account's seize rate, the rate of messages sent, or the total number of outgoing messages. Once a significant change has been detected, the ISP can proceed very similarly to how a bank deals with suspected fraud. As with any unusual spending activity, the account would be temporarily "frozen", unable to send out

more messages until input is received from the account owner. The owner would have a choice of how to proceed. If he acknowledges the messages in question, of course, this detection was a false positive and the account can be unfrozen. In this case, he is still responsible for paying for his own seized bonds. If, on the other hand, he claims no knowledge of the questionable messages, then the machine is assumed to be infected and appropriate steps are taken. In this case the ISP would cover the cost of the seized bonds. The steps taken in the case of infection will be described further in the next subsection.

Proposition 1 *If botnet detection techniques are applied, the average number of spam messages sent by an infected machine can be limited to a constant k , rather than a possibly unbounded value $m_i - m_n$.*

Proof:

Simple Bayesian probability techniques can be applied to achieve early detection with a high degree of certainty. When a machine is not infected, it is expected that its seize rate, the probability r_n of any given message having its bond seized, is low. When an infected machine is sending out spam, the number of spam messages is much greater than the number of messages sent by the legitimate account owner. This results in a drastic increase in bonds being seized. The seizure rate for an infected machine r_i would be expected to be much higher. We can at the very least assume that $r_i > r_n$.

It is possible to detect a change in seize rate very quickly. Suppose that after k consecutive bonds being seized the account is frozen. The degree of certainty in this case would be the probability that the machine is infected, and thus that the seize rate $p = r_i$, given that k messages were seized in a row. Applying Bayes' Law gives us the following:

$$Pr(\text{infected}|k \text{ seizures}) = Pr(p = r_i|k \text{ seizures}) \quad (3)$$

$$= \frac{Pr(k \text{ seizures}|p = r_i) \cdot Pr(p = r_i)}{Pr(k \text{ seizures})} \quad (4)$$

$$= \frac{r_i^k \cdot I}{r_i^k \cdot I + r_n^k \cdot (1 - I)} \quad (5)$$

Trying some sample numbers in this equation will provide some perspective. The seize rates r_n and r_i could be estimated at 0.05 and 0.95, respectively. The extremely low value for the normal rate is based on the very rare occurrence of “poor” ratings observed in reputation systems such as the one eBay employs. (Beyene *et al.*, 2008) We are assuming that each machine has an equal probability of being infected, so that the probability of a particular machine being infected is I . The probability above increases with k and with I . Even very low values for these variables, for example $k = 3$ and $I = 0.15$, results in a certainty of 99.92% when Eqn 5 is applied.

The number of spam messages sent by any single machine can now be limited to k . This gives the following new welfare equation:

$$W_b = (s + r - c_b)N \cdot m_n - N \cdot I(k)(b + c_b) \quad (6)$$

We have replaced $m_i - m_n$ with the constant k . There was previously nothing keeping the infected machine from sending any arbitrary number of messages m_i , and we expect m_n to be fairly small and bounded, so this difference could be arbitrarily large.

■

Detection could be made more sophisticated by looking for k seized messages in a particular window size, rather than just k consecutively, which would allow for normal

messages to still be intermixed. The analysis for this case would follow similar lines. Jung (2006) discusses multiple possible probabilistic approaches to detecting portscanning activities. Many of these methods are easily applicable to our detection problem here, where probe packets without responses correspond to messages whose bonds were seized.

The sheer volume of unwanted messages being sent currently poses a nontrivial problem to ISPs, as their backbone networks get increasingly more congested. This effect of detection would allow for a significant decrease in congestion, and thus provide a large payoff for the ISP. In the case of filtering, spam messages do not get delivered to the end user, but they only get dropped at the receiver's end. This means that filtering does nothing to decrease traffic over the backbone, where attention bonds are a considerable help.

5.2 Prevention

When an account gets frozen, the account's owner is alerted and given a choice. We said before that he can acknowledge the messages as ones that were sent knowingly, or he can agree that they were spam, sent without his permission. There is another subtlety to this decision that has not yet been addressed. By claiming that the messages were sent by someone else, the account owner gains the right to have the seized bonds for those messages paid for by the ISP, as part of their coverage for fraudulent charges. However, this is also an admission that his machine has had its security compromised. In this case, his ability or inclination to maintain his own security has been called into question, and the ISP (according to the user agreement we envision) at this time gains the right to remotely push down patches and maintain and monitor security for this customer. If the customer is someone who does not want to hand over this degree of control to the ISP, he is also accepting the risks of fraudulent charges onto himself, and has to pay for any seized bonds personally. This way, the party maintaining security always has an incentive to do a good

job with that maintenance, because they are likewise responsible for charges accrued when that security is compromised by a bot.

Many ISPs are already employing at least some of these measures. Some “quarantine” infected machines, giving them access only to security sites such as Microsoft Windows Update until the machines are clean. Others respond to abuse reports by helping to fix the offending machine. Given that the infrastructure is already in place to do such things, expanding those efforts should be less difficult than building them from the ground up. (van Eeten & Bauer, 2008)

Customers who are lax in patching their machine and maintaining security are likely to be the victims of botnets, and thus will in the future have their security remotely maintained. They will of course still be vulnerable to some attacks, such as zero-days, but the expectation is that they will be more secure when their patches are all properly maintained. Those computers that are maintained by the ISP are patched where they were not previously, so they are more secure than they were. The other machines not under the ISP’s control have not changed. This means that the overall security among all customer machines is higher. This increase in security across the customer base should then result in a lower rate of botnet infection I . The lower rate of infection will also decrease help desk and support costs, as previously mentioned that these costs are driven by the rate of infection among the ISP’s customer base.

Proposition 2 *If the botnet infection rate decreases by a factor of*

$$\frac{Im_n(b + c_b)}{I(m_i - m_n)c_f - (c_b - c_f)m_n}$$

then offering fraud insurance is always affordable.

Proof: In order for insurance to be affordable, the welfare W_b in the bonding case must be greater than that in the filtering case. We will assume the use of both the detection and

prevention methods which were described earlier. Given that, we can use Eqn 6 for the bonding welfare, and Eqn 1 for the filtering welfare. We assume that the infection rate has changed by some factor Δ , giving a new rate of $\frac{I}{\Delta}$ in the bonding case. This gives us the following inequality:

$$\begin{aligned}
 (s + r - c_f)N \cdot m_n - N \cdot I(m_i - m_n)c_f &< (s + r - c_b)N \cdot m_n - N \cdot \frac{I}{\Delta}(m_n)(b + c_b) \\
 (c_b - c_f)Nm_n &< N[I(m_i - m_n)c_f - \frac{I}{\Delta}m_n(b + c_b)] \\
 \frac{I}{\Delta}m_n(b + c_b) &< I(m_i - m_n)c_f - (c_b - c_f)m_n \\
 \Delta &> \frac{Im_n(b + c_b)}{I(m_i - m_n)c_f - (c_b - c_f)m_n}
 \end{aligned}$$

■

Computer expert Vint Cerf has estimated the rate of botnet infection to be as high as 20-25%, naming botnet infection a “pandemic”. (Weber, 2007) This would put the number of infected machines above 100 million. Other, more conservative estimates would still result in millions of infected machines. If m_n is 10, and m_i is 1000, and the costs b , c_b , and c_f are 5 cents, 0.01 cents, and 0.04 cents respectively, then an infection rate of 15% would only have to go to 12% before all cost of subsidising the bonds would be covered. A higher start infection rate produces a higher target delta, but also a higher target infection rate.

It was mentioned previously that the detection of bots will decrease the overall amount of traffic. Prevention of infection will cause an even further decrease, providing even more breathing room on the ISP backbone. This decrease will have a corresponding decrease in the fraction of messages which are considered spam, for similar reasons.

Corollary 1 *If the infection rate decreases by a factor of Δ then the overall spam will*

decrease by a factor of

$$\frac{\Delta \cdot m_n + I(m_i - m_n)}{m_n + I(m_i - m_n)}$$

and the overall number of email messages being sent will decrease by a factor which approaches Δ as the original spam rate approaches 1.

Proof:

The spam rate is the fraction of total messages sent that are fraudulent. Before any prevention or detection mechanisms are applied, the rate is

$$\frac{I(m_i - m_n)}{m_n + I(m_i - m_n)}$$

After the infection rate has decreased, the new spam rate will be

$$\frac{I(m_i - m_n)}{\Delta \cdot m_n + I(m_i - m_n)}$$

Assuming that I decreases, $\Delta \geq 1$. The numerators in the rates are the same, while the denominator in the second rate is strictly larger than that of the first rate. Therefore, the rate of fraudulent messages is strictly decreasing with the rate of botnet infection. The total amount of spam is decreasing by a factor of Δ . The spam rate's amount of decrease will depend on how much bigger m_i is than m_n .

The decrease in the spam rate would affect all ISPs, even those who are not implementing an ABM system. This means there is some risk that if the ABM system is too effective at decreasing spam then people may become "free riders", hoping that others will implement it so that they may enjoy the advantages without doing so themselves. Similar dilemmas result when considering measures taken to decrease the pollution of the environ-

ment.

For the decrease in traffic, similar calculations give a factor of

$$\frac{m_n + I(m_i - m_n)}{m_n + \frac{I}{\Delta}(m_i - m_n)} \quad (7)$$

for the overall number of messages being sent. If $m_i \gg m_n$, this means that the spam rate will be high. As this rate approaches 1, meaning that m_n is approaching 0, the factor above becomes Δ . This means that the worse the spam problem was originally, the more the traffic will decrease. This factor also increases as I approaches 1. This is intuitively similar, as the greater the botnet problem, the greater the spam rate as well. No matter what the original spam rate was, a decrease in botnet infection should help decrease the amount of traffic going over network backbones, thus saving ISP resources. ■

6 Advertisers

Net message values s and r were established earlier. These are average values across all sender-receiver matchings. Suppose instead that there were two classes of receivers; good receivers, who want the advertisements from a particular party, and bad receivers, which do not. Given a cost of sending a message c_s , and a value v_s to the sender if a purchase results, this gives us the following:

$$s_g = v_s - c_s$$

$$s_b = -c_s$$

$$r_g = v_r$$

$$r_b = 0$$

Proposition 3 *If ϕ is the fraction of the overall population which is interested in an advertiser's message, then under attention bond s increases from $\phi \cdot v_s - c_s$ to $v_s - c_s$, resulting in entirely new welfare-positive transactions.*

Proof:

Let us first consider our control case, the perfect filter. Assume that the advertiser initially has no way to tell which class any given receiver will fall into. This means that it only makes sense to send messages to everyone or no-one, based on whether $s = \phi \cdot v_s - c_s$ is positive or negative. That is, only advertisers for whom $\phi \cdot v_s > c_s$ will send, and they will send to everyone. This first-contact learning process will be called round 0.

Under attention bond, the dynamics change somewhat. The ISP, in the process of holding and releasing bonds, will gain information about which advertising messages were of interest to a user and which ones were not. This will provide much more detailed and accurate demographic information than most advertisers can currently obtain. Given the large amounts of money that they currently spend in order to gain inferior data, they will be willing to pay the ISP for this information.

The ISP will have to have access to this information, because they are the ones managing the transactions, and they need the ability to monitor transactions in order to detect unusual usages of the system. However, this information should not be available to the general public.

Email is not currently anonymous. Anyone who cares to gain the information can, for most accounts, see the content of emails and who is sending messages to whom. Therefore, the posting of a bond does not need to be anonymous, as it is only providing information which is already available; the knowledge that a message was sent from one account to another. When someone chooses whether to return or seize a bond, however, this releases previously unavailable information. No one currently has a way to know if a message

was of interest to someone. Now, that information can be inferred from what is done with the bond for a given message. It is possible that this could cause embarrassment for people in certain situations, so it would be a good idea to make this portion of the exchange anonymous to all but the ISP. The mechanism to achieve this would be part of the security solution for the overall system.

Assuming such a security solution can be achieved, this data can be made available to advertisers, and bring additional income to the ISP. This means that advertisers can obtain the information gained in round 0 without participating themselves. This only holds if one assumes that interest in one company indicates an interest or lack thereof in other corporations. If such is true, the advertisers who participate in round 0 will remain unchanged, but after that any advertiser for whom $v_s < c_s$ can join in.

Given the ability to target first-contact messages, the expected net message value to the sender changes from $s = \phi v_s - c_s$ to $s = v_s - c_s$, because only members of G are contacted from the beginning. This means that some advertisers will participate under attention bond, but not in the perfect filter case. This, in turn, results in entirely new welfare-positive transactions, due to the use of attention bonds. Membership in G implies that r_g is positive and the message is only sent if it makes financial sense for the sender, so the net welfare of the message is guaranteed to be positive. ■

An increase in s will result in increased welfare overall, but the ISP also benefits directly from the sales of the demographic information. This is an entirely new line of business which will not require much additional infrastructure cost to the ISP once the overall attention bond system is in place.

The new welfare-positive transactions will result in additional welfare in the advertiser's market. Let the value of this increase in welfare be V . Some amount of the additional welfare in the advertiser's market will result in turn in new value in the customer market for message receivers. The messages that a customer receives are better targeted, meaning that

their average value to that customer has increased. If v_r increases then the net value r increases as well, resulting in a higher overall welfare value in the receiver market. The exact portion of additional welfare which is duplicated from the advertiser to receiver market is determined by e_{AR} . This is the externality term describing the effects that group A has on group R , where A is the advertisers and R is the message recipients. So far, we have a term V added on to the advertiser market welfare, and a term $e_{AR} \cdot V$ added on to the receiver market. Our model of two-sided networks is adapted from the model presented in Parker & Van Alstyne (2005).

When the receiver values v_r and r increase, this will make this particular ISP more attractive to consumers, which will cause some portion of all email users to transfer membership to this ISP. The increased customer base will result in more consumers for advertisers to gather demographic information from and to initiate transactions with, making this particular portion of the advertiser market more attractive. This will result in even more advertisers participating, and another increase in the number of welfare-positive transactions taking place. The increase in advertiser welfare resulting from the new welfare on the receiver side is determined by another externality term, e_{RA} . The new welfare on the receiver side was stated to be $e_{AR} \cdot V$, making this increase $e_{RA} \cdot e_{AR} \cdot V$.

As this pattern continues, each side of the network will continue to feed into the other, resulting in the following increases (to alternating sides):

$$V \rightarrow e_{AR} \cdot V \rightarrow e_{RA} \cdot e_{AR} \cdot V \rightarrow e_{AR} \cdot e_{RA} \cdot e_{AR} \cdot V \rightarrow e_{RA} \cdot e_{AR} \cdot e_{RA} \cdot e_{AR} \cdot V \dots$$

These terms can be summed up into two simpler terms. The overall increase in the advertiser market welfare will be

$$\frac{V}{1 - e_{RA}e_{AR}}$$

and the overall increase in receiver market welfare will in turn be

$$\frac{e_{AR}V}{1 - e_{RA}e_{AR}}$$

These two equations depend on the externality terms being less than 1, which they most likely are. If for some reason they were not, the terms would not converge.

This analysis could be expanded to include such nuances as the possibility for multiple rounds of purchases from the same company, if the good for sale is consumable.

7 Discussion and Possible Extensions

The paper so far addresses the economic features of an attention bond system and its effects on botnets, but does not discuss in detail how such a system would be deployed. This includes both a detailed security solution, and a method for encouraging early adoption of the system. Attention bond systems would face an initial hurdle for adoption similar to that which was experienced by early credit card companies. Advertisers will not want to use it unless there are many consumers to be reached, and consumers will not want to switch until they see that the targeting of advertisements are indeed providing additional value. The deployment of the various elements described above will be a non-trivial problem, but one which is not within the scope of this paper. These elements have the potential, however, for producing much more surplus for the ISP, so it is reasonable to expect some up front costs in exchange for that future payoff.

The bond exchange in particular needs a strong security solution. It is an exchange of money, and as such will be a target for abuse. There is the potential that many small bond exchanges could be used to obfuscate one large transfer of money, thus providing a tool for money laundering and other criminal activities. On the other hand, if a solid security scheme is used, this mechanism could be expanded to enable legitimate payments and money transfers. This would allow ISPs to expand into other markets, competing with Paypal, Western Union, and even credit card companies. The bond exchange system would serve as a good test, allowing the system to be vetted on small payments before larger ones are enabled.

Deep packet inspection and message content analysis were mentioned before as a resource drain for ISPs, but it should be noted that they also pose an inherent risk to the privacy of the end user. ABM, by allowing messages to be vetted without being accessed, therefore helps to preserve privacy as well. This is much more in line with recent EU legislation regarding privacy and the role an ISP should have in relaying customer data. ISPs are concerned not only with compliance with such laws, but also with avoiding any liability issues regarding client data. (van Eeten & Bauer, 2008)

The botnet elements in particular have the potential for expansion. Botnet detection could feed into more general botnet research. Researchers could receive copies of newly found infections and thus stay abreast of any new trends or methods in the world of bots. As an extension of this, it is conceivable that the network traffic information which the ISP has access to could provide clues as to the source of a given infection. If an infection can be tracked to its source, this would allow even more infections to be prevented, and could potentially lead back to the actual bot farmer in time. This sort of work would be difficult to automate, and would involve holding on to much additional information, so it remains to be seen if it would be worth the effort.

The advertisement targeting information which is made available in an attention bond

system would be useful not just for email advertisements but also for webmail clients in which targeted ads are shown on the webmail page itself. This sort of opportunity would be of interest to companies like Google, which already does something similar to this in Gmail.

Finally, it should be mentioned that it is important that the bond value be something controllable by the end user, not by the ISP. Some have asked why this could not be centralized, as that would simplify the system by allowing bonds to be more uniform. However, this could potentially cause problems where net neutrality is concerned. If the ISP has control over bond values, it could set lower bonds for emails between its own customers, and higher ones from outside its own network. Such a policy would be very damaging for network neutrality. To avoid even the possibility of such an issue, the end user should keep control of the bond values.

8 Conclusions

The Attention Bond Model, at first glance, provides a means for detecting and deterring spam, and a means of detecting spam-bots that used to fly below the proverbial radar. The early detection limits the spam messages that any one bot can send out to a small constant, thus limiting the overall number of spam messages. The sheer volume of spam messages is currently a concern, and ABM will go towards alleviating that load.

The methods currently available to the ISPs for botnet detection are limited. The most common practice is to simply rely on the abuse reports sent by other ISPs and security organizations. With attention bonds, the ISPs have a means of monitoring their network for bots that involves no violations of customer privacy.

We have shown that the problems of spam and spambots are reducible, with the help of attention bonds, to a problem more akin to credit card fraud. This allows ISPs to combat

botnets using the approaches which are being used by credit card companies to combat fraud.

The measures described for preventing botnet infection, if applied, can decrease the overall rate of infection and decrease the volume of backbone traffic in another way. Other problems stemming from bots should also be to some extent ameliorated. This decrease in the incidence of bots also allows the subsidy of lost bonds by the ISP. This way, customers are not penalized if they do not know how to maintain proper computer security, and the ABM system pays for itself.

The client security maintenance needed for prevention appears expensive at first glance, but it saves the ISP money in multiple other ways, and it is only an extension of methods that they are already starting to employ.

The exchange of a bond also reveals information about the preferences and interests of the message recipient. While it is important that the systems security solution keep this data from being available to eavesdroppers, there is also an opportunity inherent there. The knowledge of a recipient's interest or disinterest in a message can allow for better, cheaper advertisement targeting. New advertisers, who could not afford to participate before, will be able to send targeted messages. This will result in entirely new welfare-positive transactions, and a new market for the ISP, in the sale of this data to advertisers. The improved targeting will increase the average recipient value of messages, which will draw new customers to the system, resulting in a two-sided network dynamic.

There are aspects of the implementation of the Attention Bond system which still need to be worked out. Security solutions regarding the exchange of bonds and the privacy of bond seizure in particular need to be addressed, but these issues are not in the scope of this paper. There are also some non-technical barriers to deployment. A business plan for bootstrapping an Attention Bond system needs to be developed as a part of future work. Credit card companies may again prove to be a good example to work from in coming up

with a plan for ABM systems. What this paper does show is that it is well worth the time and effort necessary to work out such solutions.

References

- Ayres, Ian, & Nalebuff, Barry. 2003. Why Not? How to Use Everyday Ingenuity to Solve Problems Big and Small.
- Bazeley, M. 2003. New Weapon for Spam: Bounty. April 26.
- Beyene, Y., Fauloutsos, M., Chau, Duen Horng, & Faloutsos, C. 2008. The eBay graph: How do online auction users interact? *Pages 1–6 of: Computer Communications Workshops*. INFOCOM.
- Binkley, James R., & Singh, Suresh. 2006. An algorithm for anomaly-based botnet detection. *Pages 7–7 of: SRUTI'06: Proceedings of the 2nd conference on Steps to Reducing Unwanted Traffic on the Internet*. Berkeley, CA, USA: USENIX Association.
- Cranor, Lorrie Faith, & LaMacchia, Brian A. 1998. Spam! *Communications of the ACM*, **41**(8), 74–83.
- Detica. 2006 (October 12). *BT Purchases StreamShield's Content Forensics for 'Spam Buster' System*. Press Release.
- Dwork, Cynthia, & Naor, Moni. 1993. Pricing via Processing or Combatting Junk Mail. *Pages 139–147 of: Advances in Cryptology – CRYPTO 1992*. Lecture Notes in Computer Science, no. 740. Springer-Verlag.
- Fahlman, Scott E. 2002. Selling Interrupt Rights: A Way to Control Unwanted E-Mail and Telephone Calls. *IBM Systems Journal*, **41**(4), 759–766.
- Friedman, E, & Resnick, Paul. 2001. The Social Cost of Cheap Pseudonyms. *Journal of Economics and Management Strategy*, **10**(2), 173–199.

- Goebel, Jan, & Holz, Thorsten. 2007. Rishi: identify bot contaminated hosts by IRC nickname evaluation. *Pages 8–8 of: HotBots'07: Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets*. Berkeley, CA, USA: USENIX Association.
- Graham-Cumming, John. 2004. Beating Bayesian filters. *In: MIT Spam Conference*.
- Gu, Guofei, Porras, Phillip, Yegneswaran, Vinod, Fong, Martin, & Lee, Wenke. 2007. BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation. *In: Proceedings of the 16th USENIX Security Symposium*.
- Gu, Guofei, Perdisci, Roberto, Zhang, Junjie, & Lee, Wenke. 2008a. BotMiner: clustering analysis of network traffic for protocol- and structure-independent botnet detection. *Pages 139–154 of: SS'08: Proceedings of the 17th conference on Security symposium*. Berkeley, CA, USA: USENIX Association.
- Gu, Guofei, Zhang, Junjie, & Lee, Wenke. 2008b (February). BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic. *In: Proceedings of the 15th Annual Network and Distributed System Security Symposium (NDSS'08)*.
- Hermalin, Benjamin E., & Katz, Michael L. 2004. Sender or Receiver: Who Should Pay to Exchange an Electronic Message? *RAND Journal of Economics*, **35**(3), 423–447.
- Jung, Jaeyeon. 2006 (June). *Real-Time Detection of Malicious Network Activity Using Stochastic Models*. Ph.D. thesis, MIT.
<http://nms.lcs.mit.edu/papers/thesis-final.pdf>.
- Kraut, R, Sunder, Shyam, Morris, J, Cronin, M, & Filer, D. 2003. Markets for Attention: Will Postage for Email Help? *Pages 206–215 of: ACM Conference on CSCW*.

- Krishnamurthy, Balachander. 2004. *SHRED: Spam Harassment Reduction via Economic Disincentives*. Working Paper, AT&T Research.
- Laurie, Ben, & Clayton, Richard. 2004. Proof-of-Work Proves Not to Work. *In: Workshop on Economics and Information Security*.
- Libbey, Miles. 2004. *Learning from 2003: Spamming Trends and Key Insights*. Presentation at the 2004 MIT Spam Conference. <http://spamconference.org/talks2004.html>.
- Lim, Jamus Jerome. 2008. Zombies May Mean Attention Bonds Will Not Cure Spam. *Economists' Voice: Letters*, 5(2). Article 5.
- Loder, Thede, Van Alstyne, Marshall, & Wash, Rick. 2006. An Economic Response to Unsolicited Communication. *Advances in Economic Analysis & Policy*, 6(1), 1–36. Article 2.
- McPherson, Danny, & Labovitz, Craig. 2007 (September). *Worldwide Infrastructure Security Report, Volume III*. Tech. rept. Arbor Networks. <http://www.arbornetworks.com/repor>.
- Pantel, Patrick, & Lin, Dekang. 1998. SpamCop– A Spam Classification & Organization Program. *In: Proceedings of AAAI-98 Workshop on Learning for Text Categorization*.
- Parker, Geoffrey, & Van Alstyne, Marshall. 2005. Two Sided Network Effects: A Theory of Information Product Design. *Management Science*, 51(10), 1494–1504.
- Rainie, Lee, & Fallows, Deborah. 2004 (March). *The CAN-SPAM Act has not helped most email users*. http://www.pewinternet.org/report_display.asp?r=116.

- Sahami, Mehran, Dumais, Susan, Heckerman, David, & Horvitz, Eric. 1998. A Bayesian Approach to Filtering Junk E-Mail. *In: Proceedings of AAAI-98 Workshop on Learning for Text Categorization.*
- Timothy. 2005 (May 17). *Selling Your Attention to Spammers.* Slashdot. <http://it.slashdot.org/article.pl?sid=05/05/17/1752218&from=rss>.
- Tompkins, Trevor, & Handley, Dan. 2003. Giving E-mail back to the users: Using digital signatures to solve the spam problem. *First Monday*, **8**(9). http://firstmonday.org/issues/issue8_9/tompkins/index.html.
- van Eeten, Michel J.G., & Bauer, Johannes M. 2008 (May 29). *Economics of Malware: Security Decisions, Incentives and Externalities.* Tech. rept. 2008/1. OECD Directorate for Science, Technology and Industry.
- van Zandt, Timothy. 2004. Information Overload in a Network of Targeted Communication. *RAND Journal of Economics*, **35**(3), 542–560.
- von Ahn, L., Blum, M., Hopper, N., & Langford, J. 2003. CAPTCHA: Using hard AI problems for security. *In: Proceedings of EuroCRYPT.*
- Weber, Tim. 2007 (January 25). *Criminals 'may overwhelm the web.* BBC News. <http://news.bbc.co.uk/2/hi/business/6298641.stm>.