

A Game Theoretic Model and Empirical Analysis of Spammer Strategies

Manoj Parameswaran
Santa Clara University
Santa Clara, CA 95053
manoj.pc@gmail.com

Huaxia Rui
University of Texas at Austin
Austin, TX 78712
ruihuaxia@mail.utexas.edu

Serpil Sayin^{*}
Koç University
Sariyer, Istanbul, Turkey
ssayin@ku.edu.tr

ABSTRACT

Network security problems are deteriorating worldwide, and can potentially undermine the growth of the digital economy and imperil the multitude of innovations that have been a significant driver of economic growth as well as providing increased services to individuals, businesses, and governments. The emergence of botnets as a powerful force undermining security has raised new and important issues. In particular, the difficulty of detection, elimination and prevention of botnets or spam caused thereof on an absolute scale using computing technologies alone have focused attention on studying behavior patterns of botnets and spammers, to help devise better countermeasures. This paper has two objectives; first to introduce a theoretical modeling approach to spammer behavior and derivation of the model, and second, to compare some of the derivations with data that has been collected from blocklist organizations. By making inferences about the blocklist rules, the spammer can strategize to maximize the amount of spam sent, and we find evidence of spammers using multiple strategies. The blocklist can achieve reduction of spam by investigating longer history of a node's behavior instead of focusing on detection alone. While some of the derivations seem consistent with the data there is considerable room for modification and extension of the modeling approach. The paper concludes with suggestion for the extension of the model.

1. INTRODUCTION

The technique of using blocklists to fight spam has been widely used in the industry. Blocklists serve as a convenient tool for mail service providers (MSP) to efficiently filter out spam messages. In addition, the existence of blocklists also deter spammers from sending out spam messages unscrupulously. If spammers are rational agents, they will respond to the blocklist rules optimally to maximize their benefits from sending spam messages. Hence, if the assumption of rational spammers is correct, then the data from various blocklists should reflect the behavioral characteristics of the spammers. The purpose of this paper is to investigate this through both theoretical modeling and empirical analysis.

Various streams of research have looked into analyzing

spam and blocklist data in studying spam behavior and devising improved ways to control spam. For example, [11] studies spammer behavior at network level and show that most spam arrives from a few regions of address space, and there is evidence of bot use as a key factor. [12] designs a filtering system that tracks sender patterns across multiple source addresses to track spammers who keep individual addresses below the radar and/or vary addresses.

Network researchers are increasingly recognizing the fact that spam is motivated by economic profit. [6] describes infiltration of a botnet to assess potential payoff from spam and show that even with high filtering rates, spammer has a high payoff that can run into millions of dollars a year. The spam economy is estimated by some to be worth multi-billion dollars. Research has also looked at studying spam from the spammer's point of view, for instance [10], shows how monitoring spam as it is created in infested nodes can be used to filter spam with a high degree of effectiveness. Research into spammer behavior has been mostly empirical in nature. Our approach is to supplement empirical analysis with theoretical models, in particular, dynamic models that account for behavior over time. We draw from game theory and control theory ([2]) in developing the model. The intention is to generate predictions of behavior using the model, and then to test the predictions against blocklist data, which in turn help us revise the model. In the process, we make several simplifying assumptions and introduce abstractions to represent spammers and those who defend against spam. This model is a starting point in a series of theoretical and empirical analyses intended to tie together patterns at single IP level, at botnet level and at network level that we study. As such, the focus here is limited to interaction at the level of a single source, and both empirical findings and adding richness will substantially revise the model.

Computer Science researchers have come to focus on game theory in a variety of contexts. [5] provides a brief survey. Such use has been focused on the algorithmic aspects, such as mechanism design, computational complexity and application of game theory to distributed computing problems. Economics, on the other hand, uses game theory to solve problems where multiple agents are maximizing their economic payoffs, and uses analysis to characterize optimal strategies. Where strategic behavior by two parties contending for economic gains is present with interdependence of actions, game theory proves to be a valuable tool to study behavior patterns that may emerge, and can in turn guide choices on how to influence behavior. Our use of game theory follows the economics approach, as the context of spam

^{*}Currently on sabbatical leave at University of Texas at Austin.

essentially involves economic motivation. The spam economy is substantial, and spammers are motivated by payoffs.

Game theoretic models have been applied to spam research in various contexts. [1] developed a two-player game between spammers and mail users to predict respective strategies that may help tune spam filters. [14] extended this model to account for Human Interactive Proofs in conjunction with spam filters. [13] compared the effects of different anti-spam methods, namely, filters, do-not-spam registries, and increasing cost of sending mail, on spam incidence at user inboxes and total spam volume, using game theoretic models.

The game theoretic model represents a subset of our research, focusing on characterizing spammer strategies in a simplified setting. The intention of this model is to drive empirical analysis and development of more complex models that also provide input to development of reputation rankings for providers. The development of an economic model is influenced by our view that in conjunction with technological approaches, spam should also be studied as an economic problem, using economic theory to model profit-motivated behavior and to guide empirical analysis, as well as using economic principles to develop strategies to control spam. The intention is to develop policy guidelines for deployment of technological solutions. This is in line with a long stream of research advocating that Internet policy should take cognizance of the economic aspects of the environment in guiding deployment of technologies ([4], [7]). We develop this approach further in concurrent research into designing reputation rankings and incentive mechanisms for mail service providers to motivate better control of spam, empirical analysis of data to categorize providers and prescribe appropriate policies, and study of specific events like botnet takedowns.

Our game theoretic model is differentiated by the modeling of dynamic games as well. We consider a stream of payoffs to the spammer over time rather than model single period games.

In prior work [8], we outlined the need for economic incentives to be allocated to motivate service providers in optimally deploying technological controls for security, emphasizing the reduction of attacks originating from their respective networks as against minimizing impact to their customers alone. In the context of spam, we demonstrated [9] that a certification mechanism can provide incentives that would lead to the more secure providers coalescing into a coalition that trusts each other in optimal filtering, thus making it easier to discern the degree of likelihood of a message being spam based on the provider it is coming from.

The paper is organized as follows. Section 2 presents our theoretical model. In Section 3, we present some empirical evidence. We conclude in Section 4.

2. THE MODEL

We focus on a game between two players: the spammer and the *monitor*. The spammer is in control of an IP address and wants to maximize the benefit he derives from sending out spam messages. His utility function $u(\cdot)$ is an strictly increasing and strictly concave function of the amount of spam messages he sends. More specifically, we assume $u(\cdot)$ is continuously differentiable, $u(0) = 0$, $u' > 0$, $u'' < 0$.

The other player of the game is a monitor whose role is helping out the mail service providers (MSP) in the fight

against the spammer by putting those IP addresses that send out spam messages on a blacklist. The blacklist can then be used by adopting MSPs to filter out spam messages. Not all of the MSPs use the blacklist and we assume that the weighted-proportion¹ of those MSP that do not adopt the blacklist is $\theta \in (0, 1)$.

We assume that the monitor places an IP address into the blacklist if the total volume of spam messages within any time interval of length T exceeds D .² Given the dynamic nature of the game between the spammer and the monitor, we notice that blacklist companies continuously augment this fundamental rule with more rules along the way as well as possibly alter T and D in time. In the interest of intuitive clarity, we assume that the choice of T and D fully characterizes the actions of the monitor.

2.1 Spammer's Strategy

Being listed on the blacklist has a detrimental effect on the spammer's returns because spam messages sent out by him will be dropped by those MSPs that adopt the blacklist. Still, one strategy the spammer could employ would be to simply ignore the blacklist and send spam at the maximum rate \bar{i} which will immediately cause the IP address to be listed by the monitor. The spammer's benefits in this case will be due to MSPs that do not refer to the blacklist and is given by

$$U_1 = \int_0^\infty e^{-rt} u(\theta \bar{i}) dt = \frac{u(\theta \bar{i})}{r}.$$

where r is his discount rate for future cash flows. We call this strategy the defiant strategy by the spammer.

Alternatively, the spammer could try to evade the monitor by carefully designing his spam intensity function $i(t)$, $0 < t < \infty$, which determines the rate of sending spam at any time t . If the IP address is not listed with such strategy, his utility will be $\int_0^\infty e^{-rt} u(i(t)) dt$.

Given the monitor's strategy which is determined by T and D , the spammer needs to specify $i(t)$ to so as maximize his utility while evading the monitor. Thus the the spammer's problem can be expressed as

$$\begin{aligned} \max_{i(t)} \quad & U = \int_0^\infty e^{-rt} u(i(t)) dt \\ \text{s.t.} \quad & \int_s^{s+T} i(t) dt \leq D, \forall s \in [0, \infty), \quad i(t) \geq 0. \end{aligned} \quad (1)$$

We call the resulting action the evasive strategy of the spammer. Although the defiant strategy is straightforward, the evasive strategy could be subtle. Before we state our main result, we first show that all the constraints in the above problem must bind.

LEMMA 1. *In an optimal solution to the constrained maximization problem (1), the volume constraints always bind,*

¹MSPs that serve more email accounts have larger weight than those that serve fewer email accounts.

²Since the monitor might not be able to detect all the spam messages sent out to the Internet by the spammer, D should be interpreted as the accuracy-adjusted threshold. In other words, if we assume the monitor is able to detect a proportion p of the total spam messages sent out by the spammer, then the threshold should be multiplied by p .

i.e.,

$$\int_s^{s+T} i(t)dt = D, \forall s \in [0, \infty)$$

PROOF. Suppose for some s_0 , $\int_{s_0-T}^{s_0} i(t)dt + 2\epsilon = D$ where $\epsilon > 0$. Since $\int_{s-T}^s i(t)dt$ is a continuous function of s , there exists $0 < d \ll T$ such that $\int_{s-T}^s i(t)dt + \epsilon < D$, $\forall s_0 - d \leq s \leq s_0 + d$. Let $\delta = \epsilon/d$. If we modify the intensity function to $i'(t)$ such that $i'(t) = i(t) + \delta$, $\forall s_0 - d < t \leq s_0$, $i'(t) = i(t) - \delta$, $\forall s_0 < t < s_0 + d$, and $i'(t) = i(t)$, $\forall t \notin (s_0 - d, s_0 + d)$, then all the constraints are still satisfied, but the spammer is better off because he discounts future cash flow. \square

The next result is our main theoretical foundation for studying spammers' behavior. We show that the spammer's optimal intensity function is periodic with period T . In addition, we fully characterize the spammer's intensity function.

PROPOSITION 1. *The spammer's optimal intensity function is periodic with period T , i.e.,*

$$i^*(t+T) = i^*(t), \forall t \in [0, \infty).$$

and

$$i^*(t) = v(Ce^{rt}), \forall t \in [0, T) \quad (2)$$

where $v(\cdot)$ is the inverse function of $u'(\cdot)$ and $C > 0$ is a constant such that $\int_0^T i^*(t)dt = \int_0^T v(Ce^{rt})dt = D$.

PROOF. By Lemma 1, $\int_s^{s+T} i(t)dt = D, \forall s \in [0, \infty)$. Taking derivative with respect to s on both sides yield $i(t+T) - i(t) = 0$, i.e., $i(t)$ is a periodic function with period T . Hence the spammer's problem reduces to the following optimization problem:

$$\begin{aligned} \max_{i(t)} \quad & U = \int_0^T e^{-rt} u(i(t)) dt \\ \text{s.t.} \quad & \int_0^T i(t) dt = D. \end{aligned} \quad (3)$$

If we define the state variable $x(t) = \int_0^t i(s)ds$, the above problem could be written as the following variational problem:

$$\begin{aligned} \max_{x(t)} \quad & U = \int_0^T e^{-rt} u(x'(t)) dt \\ \text{s.t.} \quad & x(0) = 0, \quad x(T) = D. \end{aligned} \quad (4)$$

The necessary condition for a continuously differentiable solution $x(t)$ is that $x(t)$ must satisfy the Euler's equation [3], i.e.,

$$F_x - \frac{d}{dt} F_{x'} = 0,$$

where $F(t, x, x') = e^{-rt} u(x'(t))$ in our problem. The above necessary condition implies

$$0 - \frac{d}{dt} (e^{-rt} u'(x'(t))) = 0,$$

$$e^{-rt} u'(x'(t)) = C,$$

where C is a constant. Denote the inverse function of $u'(\cdot)$ by $v(\cdot)$. We can write the solution as $x'(t) = v(Ce^{rt})$, i.e., $i(t) = v(Ce^{rt})$. The constant C is determined by the boundary condition $\int_0^T v(Ce^{rt})dt = D$ \square

The above result states that the spammer's intensity function will repeat itself at an interval of length T . Therefore it suffices to characterize the optimal intensity for the period $[0, T)$, which we show that depends on the spammer utility function and his discount rate as well as the parameter D set by the monitor. Intuitively, the spammer will adjust his intensity so as to utilize his allowance of D amount of spam within a period of T taking into consideration the utility he derives from his returns.

The result below states that if a spammer employs an evasive strategy, his utility will decrease as T increases, and his utility will increase as D increases. Given a T , a higher D means that the monitor is more tolerant of spam and leaves the spammer with more room for returns. Given a D , a longer T means that the amount of spam the monitor tolerates per time period is lower, leaving the spamming with a lower utility.

COROLLARY 1. *Spammer's utility from the evasive strategy is*

$$U_0 = \frac{1}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rt})) dt$$

which is decreasing in T and increasing in D (or decreasing in C equivalently).

PROOF. The utility expression follows directly from Proposition 1. The monotonicity with respect to D is obvious since U_0 is decreasing in C which is decreasing in D . To show U_0 is decreasing in T , we need to check if

$$\frac{dU_0}{dT} < 0$$

i.e.,

$$\frac{e^{-rT} u(v(Ce^{rT}))}{1 - e^{-rT}} < \frac{r e^{-rT}}{(1 - e^{-rT})^2} \int_0^T e^{-rt} u(v(Ce^{rt})) dt$$

or equivalently

$$u(v(Ce^{rT})) < \frac{r}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rt})) dt \quad (5)$$

By strict concavity, $u'(\cdot)$ is decreasing, hence, its inverse, $v(\cdot)$, is also decreasing. By strict monotonicity, $u(\cdot)$ is increasing, hence, $u(v(\cdot))$ is a decreasing function. Therefore, we have

$$\begin{aligned} & \frac{r}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rt})) dt \\ & > \frac{r}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rT})) dt \\ & = \frac{u(v(Ce^{rT}))}{1 - e^{-rT}} \int_0^T r e^{-rt} dt \\ & = u(v(Ce^{rT})). \end{aligned}$$

\square

Now that we have defined the defiant and evasive strategies as the two extreme ways of action that are possible for a spammer, below we investigate how the two strategies relate to each other.

LEMMA 2. *One necessary condition for the spammer to be indifferent between the evasive strategy and the defiant strategy is*

$$B > u(v(Ce^{rT}))$$

where B is defined as $B = u(\theta \bar{i})$

PROOF. The spammer is indifferent between the evading strategy and the defying strategy if and only if $U_0 = U_1$, i.e.,

$$\frac{1}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rt})) dt = \frac{B}{r}.$$

From the proof of Corollary 1, we know

$$B = \frac{r}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rt})) dt > u(v(Ce^{rT})).$$

□

The above result establishes a necessary condition for the indifference of the spammer between the two strategies. Below, we show that the spammer's discount rate, or his valuations of his returns over time, will make him prefer one strategy over the other.

COROLLARY 2. *Suppose the spammer is indifferent between the evasive strategy and the defiant strategy, then an increase of his discount rate r will make him strictly prefer the defiant strategy and a decrease of r will make him strictly prefer the evasive strategy.*

PROOF. We need to show

$$\frac{dU_0}{dr} < \frac{dU_1}{dr}$$

when parameters are such that $U_1 = U_0$. First, we rewrite U_0 in the following form

$$U_0 = \frac{1}{r(1 - e^{-rT})} \int_{e^{-rT}}^1 u(v(C\frac{1}{x})) dx$$

where we have used $x = e^{-rt}$ as the change of variable for the integration.

$$\begin{aligned} \frac{dU_0}{dr} &= \frac{T}{r(1 - e^{-rT})} u(v(Ce^{rT})) e^{-rT} \\ &\quad - \frac{1 - e^{-rT} + Tre^{-rT}}{r^2(1 - e^{-rT})^2} \int_{e^{-rT}}^1 u(v(C\frac{1}{x})) dx \\ &= \frac{T}{r(1 - e^{-rT})} u(v(Ce^{rT})) e^{-rT} \\ &\quad - \frac{1 - e^{-rT} + Tre^{-rT}}{r(1 - e^{-rT})} \frac{B}{r} \end{aligned}$$

Hence,

$$\frac{dU_0}{dr} < \frac{dU_1}{dr}$$

is equivalent to

$$\frac{T}{r(1 - e^{-rT})} u(v(Ce^{rT})) e^{-rT} < \frac{B}{r^2} \left(\frac{1 - e^{-rT} + Tre^{-rT}}{1 - e^{-rT}} - 1 \right) \quad (6)$$

$$\Leftrightarrow u(v(Ce^{rT})) < B$$

which is true by Lemma 2. □

This result implies that a spammer who has a high discount rate, i.e. who values current returns much more than future returns, is more likely to adopt a defiant strategy. It is possible to interpret such an approach as a myopic attitude to maximizing overall utility. On the other hand, a spammer who has a lower discount rate, meaning that he has a higher valuation of the future returns to be collected, will be more inclined to use an evasive strategy.

2.2 Monitor's Strategy

The analysis in Section 2.1 gives us a way of formalizing the spammer's possible actions as a function of the parameters that are defining the monitor's rules. While formulating her strategy and determining D and T , the monitor will take the spammer's optimal actions into consideration. As the monitor is assumed to be non-profit-seeking, her objective is to minimize the amount of spam messages released by the spammer.

Formally, the monitor's problem is

$$\min_{T \geq 0, C < 0} R(T, C) = \frac{1}{T} \int_0^T v(Ce^{rt}) dt \quad (7)$$

$$s.t. \quad \frac{1}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rt})) dt = U_1. \quad (8)$$

where $R(T, C)$ is interpreted as the average rate of spam messages sent out to the Internet and $v(Ce^{rt})$ in the monitor's objective function is the spammer's optimal intensity function based on Proposition 1. In an effort to maintain the spammer's indifference, the constraint (8) in the above formulation ensures that the spammer obtains the same utility from an evasive strategy as he would have obtained from a defiant strategy.

Obtaining analytical solution to the monitor's problem is in general very difficult and there might be no interior solution at all depending on the utility function of the spammer. However, certain characterization is possible with the general utility form. The following proposition gives us some idea how the solution might be.

PROPOSITION 2. *The upper bound of the monitor's objective function is $\theta \bar{i}$ and immediate blocking (i.e., T close to zero) is generally not optimal.*

PROOF. First we denote $\lim_{T \rightarrow 0} C = C_0$. Since the constraint holds for each feasible (T, C) , we have

$$\lim_{T \rightarrow 0} \frac{1}{1 - e^{-rT}} \int_0^T e^{-rt} u(v(Ce^{rt})) dt = \frac{u(v(C_0))}{r} = \frac{B}{r}$$

which implies $u(v(C_0)) = B$.

Notice

$$\lim_{T \rightarrow 0} R(T) = v(C_0) = u^{-1}(B) = \theta \bar{i} = R_0.$$

Now we will show that $\lim_{T \rightarrow \infty} R(T) < R_0$. By Lemma 2 we know $B > u(v(Ce^{rT}))$, $\forall T > 0$. Hence,

$$\begin{aligned} B &\geq \lim_{T \rightarrow \infty} u(v(Ce^{rT})) = u(\lim_{T \rightarrow \infty} v(Ce^{rT})) \\ &\Leftrightarrow R_0 = u^{-1}(B) \geq \lim_{T \rightarrow \infty} v(Ce^{rT}) \end{aligned}$$

Now, if $\lim_{T \rightarrow \infty} \int_0^T v(Ce^{rt}) dt < \infty$, then obviously we have $\lim_{T \rightarrow \infty} R(T) < R_0$. On the other hand, if

$$\lim_{T \rightarrow \infty} \int_0^T v(Ce^{rt}) dt = \infty,$$

then by l'Hopital's rule, we have

$$\lim_{T \rightarrow \infty} R(T) = \lim_{T \rightarrow \infty} v(Ce^{rT}) \leq R_0.$$

The above results imply that immediate blocking (i.e., T close to zero) is generally not optimal. The upper bound of $R(T)$ is hence $R_0 = \theta \bar{i}$. □

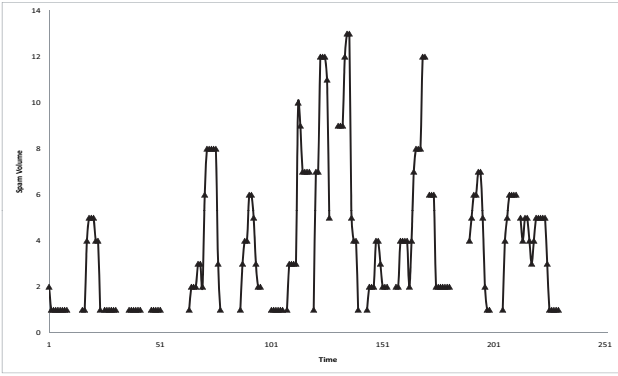


Figure 1: Low Rate Spamming Activity by IP Address Indicative of Evasive Strategy: Longer Cycles and Variable Rates

The above result has several important implications. First, it suggests that the monitor should look more into the spamming history of an IP address when deciding whether to block the IP address or not. Of course this requires more computation power and larger storage space. Hence, the monitor should trade off the benefit of large T and the cost of more resources which will lead to an interior solution. Second, the expression of the upper bound suggests the importance of the universal acceptance of the blocklist by the MSP. The smaller θ is, the smaller the upper bound is. In practice, T is always positive, hence the actual average volume of spam messages will be definitely smaller than this upper bound. Third, the result suggests that very strict blocking policy is generally not optimal because more spammers will choose the defiant strategies which causes more spam messages being sent out.

3. EXPLORATORY EMPIRICAL STUDY

For an exploratory data analysis, we built a subset from a set of daily spam listings provided by CBL which gathers data via its own spam traps. We focused on IP listings for which we have the associated volume information and we restricted our attention to IP addresses from 13 arbitrary autonomous systems that possess heterogeneous characteristics. Our data runs from July 1, 2009 to March 9, 2010 and the set contains information on 17,942 distinct IP addresses listed during this time period with 166,905 IP-day combinations. In addition to the daily spam volumes, a botnet association is provided by CBL whenever applicable. Our main goal in conducting the exploratory data analysis is to see some evidence in support of the existence of multiple strategies attributed to the spammer as described in Section 2.1. The average volume of daily spam per IP address is 135 messages, with a range of 1 to 25,558 and a standard deviation of 450.5. The variation in the spam volume across IP addresses is a clear indication of the non-uniform spamming behavior at the IP level. Figures 1- 3 depict spamming activity of some selected IP addresses over a period of 252 days. The actual dates are not labeled on the axes to avoid clutter. Periodicity of the spamming intervals can be easily observed in Figures 1 and 2 and are in line with the evasive strategy employed by a spammer. The changes in spamming rates and the length of the spamming cycles are not

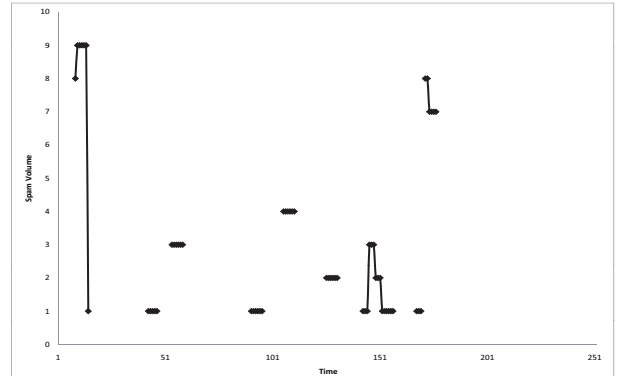


Figure 2: Low Rate Spamming Activity by IP Address Indicative of Evasive Strategy: Shorter Cycles and Uniform Rates

Table 1: Summary Statistics of Spamming Behavior by Botnets

Botnet	IP Count	Rate	Days	Total Volume
n/a	9581	94.66	7.43	7,579,416
Rustock	6074	65.99	7.19	4,357,658
Cutwail	1629	235.76	5.53	2,801,215
Cutwail2	1394	223.27	4.75	1,741,994
Bagle-cb	464	342.74	6.87	1,226,699
Mega-D	164	523.53	7.30	1,033,025
Xarvester	139	120.35	9.02	631,707
Grum	696	156.16	5.30	631,451
Bobax	52	209.37	15.27	251,904
Lethic	66	536.54	6.00	176,564
Maazben	101	266.04	4.49	132,896
Grum2	291	79.76	4.24	100,346

identical in both figures. This is to be expected, as in reality the spammer does not know the parameters D and T employed by a monitor, the spammer knows that there are several monitors with different policies, and thus formulates his strategy based on a unified perception of the monitoring environment.

Figure 3 displays the activity of an IP address that sends out spam at a rate higher than 10,000 messages per day for the most part of its active cycle, which is captured by our data as 26 days. The fact that the IP address does not appear in our data set afterwards might be due to several reasons. While it is possible that the spammer chose to stop spamming at this time, it might also be the case that the hijacked IP address was restored by cleaning up the trojans that infest the host and/or installing the security patches that eliminate vulnerabilities, thus truncating the spamming episode.

While we were able to observe similar patterns for several other IP addresses whose activity we charted, stronger inferences might be possible through a clustering study applied to a more comprehensive data set. As such a clustering study is beyond the scope of this work, below we provide some simple summary statistics that may help observe some of the variability in spamming behavior across IP addresses. In addition, we wish to see how botnets controlling the IP addresses affect the observed spamming behavior. Table 1

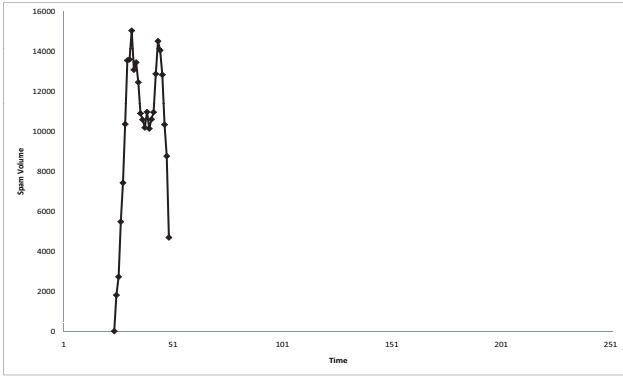


Figure 3: High Rate Spamming Activity by IP Address Indicative of Defiant Strategy

lists the botnets with a total volume of at least 100,000 messages in addition to the group of IP addresses for which no botnet information is available, which is labeled as “n/a.” The column next to the botnet names reports the number of distinct IP addresses associated with that particular botnet in our data set. The *rate* column summarizes the spam rate figures by averaging over the IP addresses associated with a botnet. We compute the spam rate of an individual IP address by dividing the total amount of spam sent over the total number of days it appears in the data set. Note that the average spam rates across botnets show a significant variability from an average of 66 messages per day for Rustock to 536 messages per day for Lethic. Note also that Rustock is the botnet with the highest total volume of spam in our data set. We also observe a variation in the number of days IP addresses spend on the blacklist on average, which is reported under the *days* column in Table 1. We believe that the variability in Table 1 is indicative of the fact that different botnets may employ different strategies to maximize their revenues within the limits of their operating realms. Characterizing the spamming strategies employed by botnets poses a further research question.

4. CONCLUSION

This is the first time that an economic model using game theory has been developed in dynamic setting to characterize spammers’ strategies. The model abstracts several complex aspects of the problem into simple representations; however, in doing so, it provides useful insight into the interaction between botnets and blocklists. More importantly, it provides the starting point for extending to strategies across multiple sources and strategies in the presence of MSPs of different security profiles. In concurrent research, we are conducting empirical analyses into spam patterns across multiple IP addresses, and working on developing a reputation system that accounts for MSPs of varying profiles. We show preliminary evidence of existence of multiple strategies by exploratory analysis on blacklist data. Results show variable strategies may be used across different IP addresses as well as different botnets. We provide insights that can guide blacklist providers in making choices about time intervals over which addresses are monitored.

There are several limitations of our current study. As mentioned earlier, our simplification of the monitor’s rule to

the choice of the parameters D and T does not adequately reflect the mode of operation of current blacklist companies. However, we believe that these are two key parameters that impact their decisions on listing IP addresses. Apart from the mathematical challenges of incorporating more complex rules into a game theoretic model that seeks to derive the spammer’s actions, it is understandable that blacklist companies would not be willing to openly discuss the complete set of rules they employ, for the obvious reason that it would provide the spammer with a tactical advantage.

Our model focuses on one spammer versus one monitor as actors whereas in reality there are several spammers and several blacklist companies. From the monitor’s perspective, we can argue that, if she desires so, the monitor may observe different categories of spammers and customize her strategies to target segments of spammers. Our analysis of monitor’s strategy provides her with guidelines to employ in order to minimize the damage that is caused by a spammer that is characterized by a particular utility function. Based on her perception of the spammer profiles, the monitor can design her rules to pursue them simultaneously. However, the problem faced by a spammer who is in interaction with multiple independent blocklists is quite different. In addition, it is not easy for the spammer to customize his strategies with respect to different monitors, mainly because this requires for the spammer to be involved in a more targeted spamming effort taking into consideration the information regarding MSP subscriptions to blocklists. Instead, we suggest that the spammer builds his strategy based on his belief of the aggregate monitoring he faces, which is derived from the partial information he gathers as he observes the reactions of the blocklists to his spamming tactics. We believe that this accounts for some of the variability we observe in spamming behavior when studied at an individual IP address level as well as the botnet level.

In addition to several reports in the literature, our own limited data analysis results on the botnets indicate that organized botnet activity is an important dimension of the spam problem. This hints to another limitation of our model which defines the spammer as in control of a single IP address because IP-level activity is what the monitor is able to detect. As botnets control several IP addresses at the same time, the spamming botnet’s actual problem involves optimizing his overall resources, which will lead to different spamming strategies when translated to single IP addresses. We believe that this presents an opportunity for further research using both modeling and empirical approaches. As the technical aspects of the botnet operations become better known, and botnet associations can be labeled more extensively by a monitor, much useful data would accrue which then can be mined to characterize the spamming behavior of the botnets. Anticipating the botnet’s behavior and tactics would in turn help the monitor in designing her monitoring rules.

We note that our exploratory empirical analysis is limited to a small subset of data and the observations might be different when the analysis is repeated at a larger scale. Extending the empirical analysis in scope along several dimensions with the goal of characterizing botnet behavior remains as future work. Such an understanding will also help develop the game theoretic model further to address strategies of botnet herders in utilizing a vast and growing collection of IP addresses.

Further work will extend the model to incorporate mail service providers at different reputation levels as players in the game. As reputations influence strategies and payoffs, they can form an incentive system; the goal of the model is to demonstrate incentives can be thus allocated to improve spam control by individual providers.

5. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. 0830852 and 0831338. We acknowledge the guidance, participation and contributions of John S. Quarterman, President and CEO, InternetPerils, Inc. in the NSF funded research on economic incentive mechanisms to reduce spam, that led to this paper. We acknowledge the contributions of Markus Sammal-lahti, Sami Sainio, Aditya Joshi, Jouni Reinikainen in gathering and analyzing data. Serpil Sayın's sabbatical leave is partially supported by a TUBITAK BİDEB grant. We thank CBL for providing the spam volume data used in this project. We acknowledge the valuable comments from Geoffrey M. Voelker, University of California, San Diego.

6. ADDITIONAL AUTHORS

Additional authors: Andrew B. Whinston (University of Texas at Austin, email: abw@uts.cc.utexas.edu).

7. REFERENCES

- [1] I. Androutsopoulos, E. F. Magirou, and D. K. Vassilakis. A game-theoretic model of spam e-mailing. In *CEAS 2005*, Stanford University, Palo Alto, CA, July 2005.
- [2] E. Dockner, S. Jorgensen, N. V. Long, and G. Sorger. *Differential Games in Economics and Management Science*. Cambridge University Press, Cambridge, UK., 2000.
- [3] I. Gelfant and S.V.Fomin. *Calculus of Variations*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963.
- [4] A. Gupta, B. Jukic, M. Parameswaran, D. Stahl, and A. Whinston. Streamlining the digital economy: How to avert a tragedy of the commons. *IEEE Internet Computing*, 1(6):38–46, 1997.
- [5] J. Y. Halpern. *Computer science and game theory: A brief survey*, *Palgrave Dictionary of Economics* (S. N. Durlauf and L. E. Blume, eds.),. Palgrave MacMillan., 2008.
- [6] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage. Spamcraft: An inside look at spam campaign orchestration. In *2nd USENIX Workshop on Large-Scale Exploits and Emergent Threats*, Boston, MA, USA, 2009.
- [7] M. Parameswaran, J. Stallaert, and A. Whinston. A market based allocation mechanism for the diffserv framework. *Decision Support Systems*, 31(3):351–361, 2001.
- [8] M. Parameswaran and A. B. Whinston. Incentive mechanisms for internet security,. In *Annals of Emerging Research in Information Assurance, Security and Privacy Services, Handbooks in Information Systems Vol. 4*, H. Raghav Rao and Shambu Uphadhyaya (eds). Emerald Publishers, 2009.
- [9] M. Parameswaran, X. Zhao, F. Fang, and A. Whinston. Reengineering the internet for better security. *IEEE Computer*, 2007.
- [10] A. Pitsillidis, K. Levchenko, V. Paxson, C. Kreibich, N. Weaver, C. Kanich, S. Savage, and G. M. Voelker. Botnet judo: Fighting spam with itself. In *proceedings of the 17th Annual Network and Distributed System Security Symposium (NDSS Symposium 2010)*, San Diego, California, 2010.
- [11] A. Ramachandran and N. Feamster. Understanding the network-level behavior of spammers. In *Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, Pisa, Italy, September 11-15, 2006.
- [12] A. Ramachandran, N. Feamster, and S. Vampala. Filtering spam with behavioral blacklisting. In *CCS '07, October 29-November 2 2007*, Alexandria, Virginia, USA.
- [13] E. Reshef and E. Solan. The effects of anti-spam methods on spam mail. In *CEAS 2006*, Mountain View, CA, July 2006.
- [14] D. K. Vassilakis, I. Androutsopoulos, and E. F. Magirou. A game-theoretic investigation of the effect of human interactive proofs on spam e-mail. In *CEAS 2007*, Mountain View, CA, July 2007.