Economic Metric To Improve Spam Detectors

Fida Gillani\textsuperscript{a}, Ehab Al-Shaer\textsuperscript{b}, Basil AsSadhan\textsuperscript{c}

\textsuperscript{a}University of North Carolina Charlotte, (sgillan4@uncc.edu)
\textsuperscript{b}University of North Carolina Charlotte, (ealshaer@uncc.edu)
\textsuperscript{c}College of Engineering King Saud University, (bsadhan@ksu.edu.sa)

\section*{Abstract}
Economic lifting has made email spam a scathing threat to the society due to its related exploits. Many spam detection schemes have been proposed employing the tendency of spam to alter the normal statistical behavior of mail traffic. Threshold tuning of these detectors is still a challenging task. Since, shooting down benign emails as spam (false positive), in pursuit of higher detection rates, can be detrimental. In this paper, we introduce a novel economic metric, based on the underlying spam economic system, to assist detectors in keeping their false positives at bay by associating detection accuracy to the spammer’s cost. Hence, the sensitivity of a detector does not need to be tuned all the way up to maximize detection, but enough to make spamming cost unbearable to the spammer. Since, spam is all about making money ultimately. We also show that the statistical features used in our spam detectors can easily differentiate spam from benign and we also show that these features are hard to evade by the spammer. Our evaluation shows the effectiveness of this approach in considerably reducing the false positives for these detectors.

\textit{Keywords:}
Spam Economics, Spam Detection, Anomaly Detection, Email Spam, Consumer Economics Theory

\section{1. Introduction}

Economic lifting has made email spam a scathing threat\textsuperscript{1} to the society due to its related exploits (like malware propagation)\textsuperscript{2, 3, 4}. Existing spam detectors\textsuperscript{5, 6, 7, 8, 9} employ the logic that the spam activity always changes the normal mail behavior. Different detectors use different statistical features to segregate spam from benign (normal). Nevertheless, threshold tuning is still a challenging task as, shooting down benign emails
as spam, in pursuit of higher detection rates, can be detrimental. Since, emails are no longer used to share just fun stories rather they have gained a mission critical status. Therefore, there exists an inherent tradeoff between the accuracy and the efficiency of such behavioral detection.

Such tradeoff and the monetary benefits of spam inspired researchers to study and model the spam economy \cite{10,11,12} and to quantify the spammer’s earning \cite{13,2}. Although, the fundamental goal of these approaches is to understand spam economy and halt its progression. But, none has developed a metric for the spam detectors to improve their efficiency. In our original work \cite{14}, we introduced a novel economic metric, based on the underlying spam economic system, to assist detectors reduce their false positives by associating detection accuracy to the spammer’s cost. Hence, the sensitivity of a detector does not need to be tuned all the way up to maximize detection, but enough to make spamming cost unbearable to the spammer. As shown in the Figure 1 we want to assist detectors finding this sweet spot to defeat spam. A sweet spot represents a threshold where detector is causing enough increase in the spamming cost of the spammer (or enough decrease in the profit of the spammer) with least false positives such that the spamming activity becomes useless to the spammer.

The first contribution of the original paper \cite{14} was in identifying 4 effective statistical mail traffic features that can distinguish spam from benign. In this extended version, we have considered 6 more features to rigorously test all available features presented in the existing literature \cite{6,5,7,8}. Furthermore, we want to assure that the final features must be hard to evade by the spammer. Therefore, we have also added evasion analysis of all these features in this extended version. We want to use best features available in our economic modeling. We use $K$ directed divergence \cite{15} measure to analyze the discerning capacity of these features. We perform this analysis
on our own dataset that is comprised of around 75,000 benign emails and around 3 million spam emails.

Our analysis reveals four features to stand out among all: (1) inter-departure time (IDT), which is the time between two consecutive emails, (2) emails per recipient (EPR), which is the number of emails sent to a recipient, (3) email size (ES), which gives the average email size, and (4) distribution of new recipients (DNR), which provides the frequency of new recipients appearing in a time window of email inspection. Afterwards, we benchmark the performance of these features using ROC [16] curves to establish a baseline to later test the impact of our economic metric.

The second contribution of the original paper was in developing a spam economic model to quantify the spammer’s utility associated to a spam activity. We use the classical consumer theory of economics to model spam economy, where spammer acts as a consumer looking to buy a product (commodity) that could maximize his/her utility. To define commodity, we have used all the parameters that the spammer would look for in renting a botnet. In this extended version, we describe the intuition of these parameters in detail. In our economic model, we assume a rational spammer behavior. According to which, a spammer will choose a commodity that would yield maximum utility. This intuitive assumption is largely used in the existing literature [11, 12].

The cost of generating a spam activity is calculated from the price quotes available in the current botnet market [17, 18]. In actual, detectors force spammer to invest more by either increasing the duration of the spam activity or using more resources (bots) without detectors knowing that. We want to develop this insight into the detectors to look for both accuracy and spamming cost. For this purpose, we use the statistical features to constrain our economic model. Now, detectors adjust these statistical features to gain accuracy which in-turn constrain the spam economic model (utility). This reduced utility forces the spammer to add cost (resources or time) which we calculate and provide to the detectors. We repeat our spam detector benchmarks to observe this increase of spamming cost. In our evaluation, we map these accuracy and cost results together to show the final improvement in the entire detection process.

We structure the remainder of this paper as follows: in Section 2 we establish the novelty of our approach through literature review. We explain the feature selection mechanism and evasion analysis of all features in Section 3 followed by the performance benchmarks of the selected features in Section 4. The discussion of the economic model is presented in Section 5. The Section 7 provides the performance benchmark of the detector with the
spammer’s cost. In the end the Section 5 provides the conclusion and future directions of our work.

2. Related Work

The purpose of this related work is to show that all spam detection techniques focus only on the behavioral divergence aspect of the problem disregarding the economic aspect completely. On the other extreme, all spam economic based studies try to understand and quantify spam economy without introducing any metric for detectors to exploit these findings. As per our understanding this is the first study that is proposing an economic metric based on spam economic model to assist detectors in tuning their thresholds.

In spam detection, some studies suggested intervention from the service providers to stop dissemination of large volumes of spam. For example, the study [5] proposed proactive blacklisting of spammer’s domain through registrar monitoring and domain registration frequency to dampen spam and another study [19] proposed ISPs to monitor the involvement of different IPs in spamming to filter their traffic. These studies do not diminish the end point remedies to cope spam problem. An end point based spam detection technique [9] used entropy to measure effectiveness of different statistical features of email traffic to differentiate spam from benign without considering spam economics. In [8], a framework called AutoRE was presented to filter out any legitimate URLs and focused on the URL that the spammer wants his victims to click on to buy his merchandise or download his malware. Using their signature method, they were able to identify botnet membership and determine which bots were used in the various spam campaigns. The work in [6] focused on the network properties of spam and showed that network-level characteristics of spam are sufficiently different than those of legitimate emails. Then another work in [7] detected spam from email server logs by measuring the change in the mail behavior of a source over time. All of these studies were very effective bot/spam detectors but without any concern to the underlying economic model.

On the economic front, the economic study [12] proposed an abstract economic model of botnet usage for DDoS attack from both bot botmaster and spammer’s perspective. They introduced the concept of honeypots (fake bots) to increase the probability of failure for the attacker. However, the authors do not associate their model to any parameters used by their filters. Some other very prominent studies [13, 2, 20, 11] rigorously analyzed the spam economics using empirical measurements based approach. They
Table 1: Variables used in SPAM Economic Model

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>Set of commodities</td>
</tr>
<tr>
<td>$c_i$</td>
<td>$i^{th}$ commodity</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>Bot bandwidth (bits/sec)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Duration of bot availability</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Average spam email size</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of bots in a botnet</td>
</tr>
<tr>
<td>$p_i$</td>
<td>Price of $i^{th}$ commodity</td>
</tr>
<tr>
<td>$b_i$</td>
<td>$i^{th}$ consumption bundle</td>
</tr>
<tr>
<td>$B'$</td>
<td>Set of affordable consumption bundle</td>
</tr>
<tr>
<td>$f_i$</td>
<td>Failure probability of $i^{th}$ commodity</td>
</tr>
<tr>
<td>$u(.)$</td>
<td>Utility function</td>
</tr>
<tr>
<td>$\mu(idt)$</td>
<td>Mean value of IDT</td>
</tr>
<tr>
<td>$W$</td>
<td>Wealth of the spammer</td>
</tr>
</tbody>
</table>

have tried to quantify the spam revenue by analyzing the spam market for months, but they did not establish any connection with the amount of spam activity required for such revenue. They provided an estimate that it requires almost 10 million spam emails to get a positive response, though. A similar study [21] provided microeconomic analyses of ecrime to develop a set of hypotheses to predict potential participating crowd. Another study [22] provided an abstract model that described the impact of reducing target density for the spammer but it did not provide any metric for the detectors to actually reduce this target density.

3. Feature Selection For Spam Detection

We want to identify the statistical features from the existing literature that could effectively throttle the spam activity. The effectiveness of any feature depends upon its capability to differentiate spam from the benign emails and its ability to make evasion harder for the spammer. We have used our own collected dataset to test the feasibility of different mail traffic statistical features. In the following, we explain our dataset and feature selection process along with evasion analysis followed by the selected feature and its performance benchmark.

3.1. Evaluation Dataset

To collect a comprehensive source-based spam mail dataset, we designed experiments to attract real spamming malware. Multiple computers were setup with Windows XP SP1 and exposed to the Internet without any protection. To prevent spam from being released onto the Internet, a mail
server and firewall were put in place and any outbound SMTP packet was routed automatically to the ‘sinkhole’ mail server where it was captured and written to a file as shown in Figure 2. The mail server was setup to respond to any SMTP connect request from the infected computers. Masters received replies from what they thought were the relay servers, but they were actually crafted reply packets from the controlled mail server using the relay server’s IP address. As far as the bots knew, their emails were being sent to millions of relay servers across the Internet.

Within 24 hours of being setup, the unprotected computers were infected with multiple pieces of malware, some of which had already communicated with their ‘Command and Control’ server and had begun sending spam. These machines were monitored for two months, as multiple spam campaigns were captured before the spammer discovered the trap and decided to terminate the bots. There were over 3 million spam emails collected from bots in home and business machines.

To perform a comparative evaluation of spam with benign emails, we also collected anonymous benign email records from 66 real-life users who voluntarily participated in this experiment. A custom utility was designed which could extract headers of sent emails through a POP/IMAP web email account. This utility with associated information of the experiment were uploaded on a website and users were invited to participate in the study by running the utility and sharing their sent email data with us. The data extracted from each benign and spam email record included the sender’s address, recipients’ addresses, date, time, and the size in bytes. Participating users were also asked to provide some basic information about themselves which could allow us to factor in sociological patterns in our study. Table 2 lists some information about the benign email data. Users that participated in our study included undergraduate and graduate students, faculty mem-
It can be seen from Table 2 that in our dataset industry professionals and students/researchers generate the smallest volumes of data (a little more than 700 emails per user). We believe that these users were using their official email addresses only for official emails, while carrying out the bulk of their email communication on personal email account. As expected, faculty members and management personal generate the largest volume of per user data (an average of 4265 emails per user) because the bulk of their emails are communicated on their official addresses. Home users had on-average 976 emails per user.

3.2. Measuring Information Divergence of Traffic Features

All of the tested features are stochastic in nature and each feature is considered a random variable. For each we compute two types of histograms using samples from the spam and the benign datasets, respectively. These samples represent non-overlapping subsets of each dataset. By normalizing these histograms, we obtain the Probability Mass Functions (PMFs) of these random variables.

In order to calculate information divergence, we have two measures entropy and K directed divergence [15]. We did not use entropy, because entropy-based measures can be evaded easily since entropy is an average measure that does not take the individual symbols’ values into consideration. More specifically, a spammer can generate a traffic feature PMF which is widely different from a benign source’s PMF, yet yielding the same entropy as the benign PMF. For instance, consider that we are operating on the inter-departure time feature PMF. A benign source generates 3 emails,
Figure 3: Figure (a) shows that spam email traffic uses small IDT values and exhibit less randomness which clearly diverge these from benign email traffic. This is observable by the wide gap in the figure. Similarly, Figure (b) shows that normal users tend to send more emails to same persons and spammer cannot afford that therefore EPR parameter shows significant divergence of spam from benign emails.

The first two emails have an inter-departure time difference of 1 hour, while the third email departs 2 hours after the second email. A spambot, on the other hand, generates 1000 emails with a time difference of exactly 0.5 seconds, and another 1000 emails with a time difference of 1 second. After normalization, the PMFs of the two sources will both be Bernoulli random variables with parameter 0.5. As a consequence, entropies of the two distributions will be identical. To cater for symbol values during information quantification, we resorted to the $K$ directed divergence [15] measure, a variant of K-L divergence [15]. For two PMFs $a$ and $b$ of a discrete random variable $Z$, $K$ directed divergence is defined as:

$$K (a||b) = \sum_{j \in \Lambda} a(j) \log_2 \left( \frac{a(j)}{a(j) + b(j)} \right),$$  

(1)

where $\Lambda$ is the image of the random variable. We use 10% of benign email traffic to generate a baseline benign (normal) distribution $a$ and then we use each sample as a test distribution $b$. Divergence between $a$ and $b$ for benign and spam email sources is computed using Equation (1). We use this divergence measure to select appropriate features to be used further in our evaluation.

3.2.1. Inter-Departure Time (IDT)

The IDT provides the time between two consecutive emails. A Spammer would want to send maximum number of emails in a short span of time. In our experiments, the IDT PMFs are computed as the number of times an
The IDT value of a particular email was seen divided by the total number of emails sent so far. The resultant divergence of benign emails (home and business) stayed around 0.1 throughout all experiments, representing the consistency of IDT behavior, as showed in Figure 3(a). The absolute value of IDT remains well above 10 seconds in all samples. Whereas, the spammer tries to maximize the number of emails sent by randomly mixing different IDT values. That is why the divergence value in spam dataset stayed around 0.6 in all experiments.

The diverging behavior of IDT across benign and spam traces makes it an attractive choice for spam detection.

Evasion Analysis: The inter-departure time detection can be evaded by a spammer who randomizes the time between emails. The time between every email would have to vary to hide behind the inter-departure time pattern of benign email sources. Thus a spammer attempting to evade inter-departure time detection would have to increase the number of distinct inter-departure time values to be almost equal to that of benign sources. In our benign email dataset, 99% of inter-departure values are unique. In order to undetected, the spammer sources must send 99% of spam emails with unique inter-departure time values.

Evasion Effect: The spammers we tracked during our investigation had a consistent inter-departure time during each spam campaign. The time period between emails was less than 10 second, which allowed the bots to send out millions of spam in just a week time. If a spammer changed his bots to send out spam with a long inter-departure times matching those of benign sources, then his spam volume would be reduced roughly by more than 99%. Over the 36 hour spam campaign that we recorded, the total spam volume would have been reduced from 3 million to 510 emails.

Anti-Evasion - Mechanism and Analysis: From our research, we found the average response rate from a spam campaign is around 0.000001, or roughly 3500 responses for every 350 million spam messages. Reducing the output volume of each bot by 99% would decrease the profit of the spammer by the same percentage. If he maintained the same botnet rental period and every other variable mentioned in Section 5 remained the same, his response rate would decrease to 0.0000001% which would cause his profits to only cover less than 1% of his expenses. Based on our calculations, we found that any response rate below 0.0000043% will create a loss for the spammer.

In actual, we have used IDT bins of one second each.
Figure 4: Figure (a) shows that as spammer, mostly, uses templates of comparable sizes therefore the spam mail does not show significant divergence as compared to most benign mails. On the other hand, as size of benign emails do not depend on each other therefore, they show significant divergence. Similarly, Figure (b) exhibits that spammer tries to reach more diverse recipients as compared to normal user. But both of these parameters are weak as compared to IDT and EPR.

3.2.2. Emails Per Recipient (EPR)

A normal user tends to send most of his emails to only a group of recipients, to whom he/she is connected socially or professionally. On the other hand, a spammer avoids sending too many emails to a subset of recipients to avoid being blacklisted. In order to calculate the divergence for this feature, we counted the number of emails that each unique recipient received. Figure 3(b) shows that the divergence of the number of emails per recipient for benign email sources is much lower, about 0.02 than the divergence of the spam email, about 0.25 where PMFs in Equation 1 are computed as the number of users who received $j$ number of emails divided by the total recipients so far. This shows that the spammer has to vary his/her victims a lot to be more effective in reaching out a bigger audience. In the spam sources, the number of emails per recipient varied from 1 to 4, while the recipients of benign sources received anywhere from 1 to 34 pieces of email over a sampling time interval.

Evasion Analysis: Spammers attempt to reach as many unique email addresses as possible during their spam campaigns, which creates a high divergence value when measuring the number of emails per recipient divergence. Evading this detection technique would require a random number of emails being sent to each recipient. If we assume that the benign sources of email that we tracked are an average of all benign sources, then a spammer would have to randomly send between 1 and 34 pieces of email to every email address in order to blend in with benign email patterns.

Evasion Effect: The spam campaigns we monitored showed each unique email address receiving between 1 and 4 pieces of email, with an average of
Figure 5: Figure (a) shows that the spammer does not significantly change the domain names (e.g., yahoo, gmail, etc.,) of recipients unlike normal users and show less randomness. Therefore, this parameter is not suitable for differentiating spam from benign. Figure (b) shows that with more bots, a small number of spam emails per recipients are required to achieve same amount of spam activity.

3.34. In the subset of emails observed, the benign source closest to the spammer’s sending pattern sent 2,559 more emails to existing recipients than the spam source; this calculation is by noting that an email is copied \( n \) times by mail servers, where \( n \) is the number of unique recipients. In order for the spammer to disguise himself as a benign source he would have to match that rate per recipient. During a spam campaign in which 1 million unique email addresses are used with a total of 3.34 million emails sent, a spammer would have to send an additional 8.1 million emails to the same 1 million addresses to hide from our detector. That corresponds to an additional 2.42 emails to every recipient. A spammer could include multiple email addresses on each email to reduce that number, but he would have to do so picking email addresses at random otherwise his divergence level would remain high and would lead to detection. **Anti-Evasion - Mechanism and Analysis:** If we safely assume a spammer only has a set number of products to sell, and a person who receives multiple spam emails for the same product would not buy more than one, then the additional emails sent to disguise his sending patterns would not provide additional profit. The increase in emails would affect his expenses, since more botnet rental time would be required to send out the additional emails. Knowing that a spammer would want to minimize the number of these useless emails, he would attempt to combine as many addresses as he could on each email. Most email servers have a maximum number of addresses configured by the administrator, or has a default maximum already set. Most ISP’s we investigated set a maximum of 100 recipients per email and their reason for this number was to prevent spam [23]. Any email that contains more than the maximum allowed number of recipients is automatically discarded. If we use 99 as the maximum number of addresses per email, then a spammer
would have to send 294,545 additional emails for every 1 million emails he sent during his spam campaign, with each non-unique recipient included on an average of 2.42 of those emails. This 29% increase in emails would require an increase of 29% in the botnet rental time period and a 29% increase in rental costs. A spammer could evade this detection with a lot of work organizing his email address list and if he is willing to give up 39% of his profits due to the increased number of emails required.

3.2.3. Receiver Domain Name Distribution (RDN)

Our intuition indicates that spam programs should send a few emails per domain to avoid blacklisting. This is unlike benign emails, particularly business emails, which might include significant repetition in domain names due to frequent email exchanges within the same enterprise or domain. A ‘full domain’ is considered to be the data after the “@” symbol in the recipient’s email address. We calculate the PMF of RDN by adding a new recipient domain name to an IP’s recipient list for spam and benign email traffic over different sampling intervals. Assuming \( x \) is the number of unique (i.e. new) domains and \( y \) is the number of total emails sent during a sampling interval, we can calculate the probability of \( x \) as \( p(x) = x/y \). The results are shown in Figure 5(a). Unlike spam emails, this figure clearly shows that benign email traffic exhibits high probability values; i.e., a significant portion of the benign emails are destined to a small subset of domain names.  

**Evasion Analysis:** The RDN detector is probably the easiest for some spammers to avoid, but only if they spend time sorting their victim’s email addresses or using a reflection spamming method. Most spammers purchase their victims’ email addresses as described in Section 5.1.3. The millions of addresses are usually from many different domains, which would require the spammer to sort them before passing them out to all of the bots.

Another evasion tactic would be to perform a reflection spam campaign. A spammer would set up one or more new email accounts on a trusted online email site, such as Gmail, Yahoo or MSN. An auto-reply message would be set up on each account, which would contain a spam advertisement. The bots would all be directed to send emails to the accounts, with a reply email address set to the victim. In this way, the bots would appear to send to the same domain millions of times.  

**Evasion Effect:** Both of the evasion techniques described above would require time from the spammer. In the first case, it would require him to sort all of the millions of email address before sending them out to all of the bots. In the second case, he would have to set up one or more separate accounts and would risk exposure due to the heavy load of emails arriving for a handful of users.
Figure 6: Figure (a) shows that if we observe more emails (benign and spam) together than the difference of new recipients tend to widen as benign users would resend emails to same group of recipients as oppose to spammer. Figure (b) shows that spammer would use large size templates to fool spam filters but a smaller more intelligent email can be used to bypass filters. Therefore, ES can be easily evaded.

**Anti-Evasion - Mechanism and Analysis:** While both methods will evade this detector, they would require a change of tactic by the spammer. He would effectively be ‘cornered’, unable to send spam in any other way, and other techniques including filters and blacklists could help identify and shut down the spam source.

### 3.2.4. Distribution of New Recipients (DNR)

Spam detectors normally exploit the weakness that the spammer tends to reach out a diverse group of recipients within a small time window than a benign user. Therefore, the distribution of new recipients in any sampling window (total number of new email addresses not seen before) is an intriguing divergence feature. This coincides with the fact that most benign users have a dedicated pool of email addresses belongs to family, friends or co-workers and they end up sending emails to the same group again and again. To calculate DNR, we kept the record of email addresses appeared in shorter sampling window and did not consider it as a new email address if it is also appeared in a the large sampling window. For each sampling window, we count the frequency of new email addresses (for smallest sample as there is was no record so all distinct email addresses are considered as new) and then divide it by total distinct email addresses seen in that sampling window to calculate the PMF of DNR. Figure 4(b) shows the divergence measure of DNR in spam and benign traces. The divergence value for spam stays steady at 0.55, while its value stayed at 0.4 for business users and 0.36 for home users. Business users had a high rate of new recipients, which is intuitive as professionally a user is connected to a bigger circle than socially.

**Evasion Analysis:** One signature of a spammer is the large number
Recipients
Emails Over Time
Business Home Spam RDN

Figure 7: Spammer does not use large lists of recipients to reach out because it could trigger filtering whereas benign users use large lists e.g., student mailing list.

of unique email addresses being used over a short period of time. The DNR detection technique uses that signature, but with time and money the spammer can avoid being detected. A spammer might think he can avoid detection by leaving a large amount of time between emails so that the detector will ‘forget’ about the earlier recipients he used. Since this detector ignores time, and instead uses previous emails, this evasion tactic will not work. The only way this detector can be avoided is to spread out emails over a very large number of bots. Each bot would only send a very small number of emails which would be too small to form a definite new recipient rate pattern.

Evasion Effect: The number of spam emails sent to each recipient in a spam campaign would vary depending on the number of products the spammer has to sell, or how many variations of spam email templates he is using. Spammers will sometimes send several emails out for the same product to the same address, but using different words or pictures hoping that at least one will get through all of the spam filters. Using the average, we found in our experiment of 3.34 emails per recipient, each bot would only be able to send around 40 emails before the pattern is recognized by the detector—this amount can vary with detector sensitivity. Figure 6(a) shows the difference in rates between a spammer and a benign source over 50 emails. After 40 emails, the difference between the two rates has increased to 0.28 and has shown steady and consistent growth allowing for easy identification by the detector. Anti-Evasion - Mechanism and Analysis: If a spammer limits each bot to 40 emails with an average of 3.34 emails per recipient, then each bot would only be able to send to 12 unique email addresses. In addition, the spammer would require a botnet of over 8.75 million bots to send out 350 million emails, the size of a typical spam campaign. Figure 5(b) shows the number of bots that would required to send out 350 million emails if each bot is limited to a set number of emails. The shaded region in the
graph shows the area that can be detected as spam by this technique. As mentioned previously, the rate calculated by this detector is the number of unique email addresses divided by the total number of emails sent. Therefore, a spammer could send out a few emails, then wait for a benign email source on the bot to send out the same number of emails. This would allow the spammer to evade this detector for a longer period of time, but it would also significantly increase his botnet rental time and would require him to change the bot code to detect and count benign emails being sent. To complicate that evasion tactic, many victim computer owners do not use a SMTP mail program, so any email tracked by the detector for that IP address would be from the spammer. The spammer would be limited to one spam campaign of 40 emails on each bot, or he would have to reuse the same 12 email addresses on each bot for every future campaign.

3.3. Measuring Inconsistency

In this section, we present a number of features that show significant inconsistencies in email headers of spam traffic as compared with the benign traffic. These features are very indicative and easy to check. However, unlike the previous measures, they can be easily evaded by spammers.

3.3.1. Inconsistent Parameters

**Inconsistent HELO-Domain:** After the TCP handshake, the first call from the sender to the mail server is ‘HELO (or EHLO) domain’. The domain string informs the mail server in which domain the sender is located, but it is not reliable since the value is not validated, can be easily spoofed or entirely left out. We found that the spammers who controlled our computers routinely altered the domain name, while our benign email user group used the same domain consistently.

In our data, 87% of the spam emails had a unique domain sent in the first SMTP packet. Of those, 83% had domain names set to a unique 8-digit hexadecimal number. The remaining 17% used the same domain as the sender’s spoofed email address.

**Inconsistent Mail Server:** Another element that frequently changed in the spam that we captured was the mail or relay server that the bot connected to before sending each email. Some of the bots had a collection of known open relay servers that they looped through so that not all of their spam appeared to come from one location. Benign email users are likely to use the same mail server, or a small collection of mail servers, provided by their ISP or online mail service.
In the spam that was captured, 49% used a unique mail/relay server. The other 51% were duplicates of mail servers used previously. This equates to millions of different mail/relay servers being used by our infected computers, since an average of only two spam emails went to any one mail/relay server. If a bot used a victim’s default mail server, he would be in danger of being identified and shutdown by the ISP, or blacklisted by the recipient’s spam filters. The fewer mail/relay servers that are used, the higher the chances that those servers will be blacklisted and become unusable to the spammer. That dependence on hiding by spreading out their load among many servers leads to a high divergence from the benign sources which can easily be identified.

**Inconsistent Sender Email Address**: Previous studies have shown that the sender’s address on spam emails is rarely used more than once in an attempt to avoid address-based spam blacklists. Those studies have found the percentage of unique email addresses is over 93% [24], while our experiment produced a unique rate of over 97%. Constantly changing email addresses can be easily detected and the source can be flagged as a possible bot. As in the two previous differences, our benign user group consistently used the same sender address through all of their emails that were captured. This difference was tracked using the same technique as mentioned previously where the detection program uses the consistency of repeats to identify the valid addresses. To evade this detection method, spammers could use the same sender’s address for each campaign, but would have to make the address vary between bots and minimize the number of emails. If they did not vary the address, then it might be added to a global blacklist before all of the bots’ emails arrived at their destinations.

### 3.3.2. Evasion Analysis of Inconsistent Parameters

Evasion Analysis: Inconsistency is what makes it hard to always identify spam and the spam source. Each part of the spamming process changes constantly, from the text in the spam message, to the mail server used to send the message, to the recipient and sender’s email addresses. When the inconsistency of spam meets the consistency of a benign email, it is very easy for detectors to separate and identify the two sources. There are three detectors in our program that look for inconsistencies where none should exist. In this case, the consistent values should be the HELO domain, the mail server name, and the sender’s email address. For a benign source, all three of these are constant since they are part of the sending mail program. A spammer could also keep these values constant for each bot and effectively
evade this detector.

Evasion Effect: A normal mail program does not ask for the sender’s email address or ask which domain should be used in the HELO handshake with the mail server. Spammers have to change these values constantly or recipient mail servers will be able to identify and filter them based on these fields. If a spammer does not change these values, they will avoid our detector, but will be identified by the recipient’s mail server and these three fields can be used as filter conditions in the future. Another possible way to evade the detector and the recipient’s blacklists is to keep the values constant and only send a few messages to different domains. This would cover up the inconsistencies and avoid being blacklisted since the recipient’s mail server would receive only a few emails. However, such a strategy costs the spammer since each bot would be limited to around 50 to 100 emails, and all emails would have to be sorted by domain such that no one bot received more than a couple for each domain.

Anti-Evasion - Mechanism and Analysis: Botmasters have learned that their number one asset is a bot. Bots are what attract botnet rental customers who bring in the money. It is imperative that bots remain undetected so that they can remain in the botnet as long as possible. Inconsistent values have helped bots go undetected in the past, which will make it hard for them to alter their processes. Limiting the bots to a small number of emails will drive up botnet rental costs and will cause botnets to be larger since more bots are required to send out the same number of spam messages. As shown in Figure 5(b), millions of bots would be required to maintain the same amount of spam if each bot is limited to a certain number of emails. Avoiding the three consistency detectors is possible, but will cause significant headache for the spammer and will also lead to possible loss of bots.

3.4. Some Less Robust Features

The three features mentioned henceforth may be able to identify some spam sources, but they did not show a clear enough difference between spam and benign email to add them to our list of spam-spotting tools. We have included them in this paper because it is possible to use these features in different ways or in conjunction with other tests.

3.4.1. Email Size

Due to spam filters, spammers typically do not send the same exact email to all of their victims. We, however, observed that the size of every
Table 3: Divergence of Recipients per Email

<table>
<thead>
<tr>
<th>Source</th>
<th>Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home</td>
<td>0.61</td>
</tr>
<tr>
<td>Business</td>
<td>0.55</td>
</tr>
<tr>
<td>Spam</td>
<td>0.68</td>
</tr>
</tbody>
</table>

email does not vary by that much due to ‘templates’. Our spammers tended to have a handful of email templates to work from and small modifications were made to each email to make it unique. From our data it appears that a few random words were changed each time, but the overall size is consistent. These basic templates varied in size, and most bots varied the template used by each consecutive email. Figure 4(a) shows the use of one template by a bot, compared to the varying size of both types of benign email. All of the templates used by our bots were larger than 1000 bytes, while our benign sources were frequently under that amount.

The average size of a captured spam message was around twice the average of both types of benign email, but in our results there were wide divergence fluctuations in the benign emails as showed in Figure 6(b). The divergence results showed in Figure 4(a) shows the potential of this parameter in best case scenario but mostly divergence values were very close in the benign and spam emails. While this test would probably identify some spam sources, the bots that we monitored used enough templates to successfully hide inside of a benign email stream.

3.4.2. Number of Recipients

Our original belief was that spammers would try and send out as many emails to as many people in the shortest amount of time. While most of our spammers had high email-to-time ratios, the number of recipients on each email was low. The average number of recipients on spam emails was just 2.2, while office and social benign emails averaged 2.4 and 1.59 respectively. Although these numbers look comparable, Figure 7 shows how benign emails have wide variations, while spam has a set lower and upper bound due to its programmatic source. Divergence values shown in Table 3 are too close to separate spam from benign. This difference was left out of the detection program, even though it requires more study and could possibly work for other spam programs that were not included in this experiment.

3.4.3. Trusted Inter-Recipient Weight

Through our experiment with benign email data, we found that not only was there a limited number of unique recipients for each source, but also
there were subgroups contained within each recipient list. Each subgroup contains a list of recipients who have been included on the same emails. These subgroups seemed to be defined by the social network of the email sender, such as family, friends, work, etc. We found that a recipient in one subgroup was not included in emails sent to recipients in another subgroup. This information is useful if you anticipate a spammer reading through a victim’s address book and limiting his spam to those people to avoid detection. Our initial results proved that these groups do exist, but due to the large amount of work that could be done in this area, it was decided this difference could be examined in a future study.

3.5. Discussion of Selected Statistical Features

Our rigorous analysis of the statistical features shows that both inter-departure time (IDT) and emails per recipients (EPR) proved to be very good candidates for an effective spam detector. Similarly, out of other features, three are comparable in their effectiveness and can be evaded by the attacker with some efforts through increased expenses in spamming activity i.e., receiver domain name distribution (RDN), distribution of new recipients (DNR) and email size (ES). We selected DNR and ES for further analysis to benchmark their performance.

4. Performance Benchmarks on Selected Features

In order to know when to stop increasing the sensitivity of a detector, when need to know how these detectors perform in first place. Therefore, considering each statistical feature as a separate detector, we tested the efficiency of each of these detectors in terms ROC curves that has become a de facto standard for intrusion detection system evaluation [16]. A detector
can classify an email into one of four categories: 1) True Positive (TP), successfully declared as spam; 2) True Negative (TN), successfully declared as benign; 3) False Positive (FP), wrongly declared as spam; 4) False Negative (FN), wrongly declared as benign. The ROC based evaluation uses two metrics, True Positive Rate (TPR) to check how accurate a detector is to detect anomaly (spam) and False Positive Rate (FPR) to check the collateral damage caused by the detector in wrongly accusing benign as anomaly (SPAM). The formula we have used for TPR is as, \( TPR = \frac{TP}{TP+FN} \), which describes the ratio of successfully detected spam and for FPR is as, \( FPR = \frac{FP}{FP+TN} \), which describes of ratio of wrongly accused among all benign emails.

Both of the datasets, benign and spam, are collected separately but during the same time frame. To test these detectors we use three weeks of data from each dataset and merge these using the time stamps of each email header. One important thing to remember here, we are not solving the problem of spam detection. We are simply selecting the right feature from the existing literature and improving their tuning. Therefore, we are using the simplest way these features can be used to detect spam. The rest of the detection experiment detail with respect to each feature is presented as follows.

4.1. Performance Benchmark of IDT

In our experiments, we have used a bin size of one minute. We calculate the average IDT of each bin if the value is below the threshold then we classify it as spam, otherwise benign. These threshold values are plotted on the X-axis of the graph in Figure 8(a), which are varied from 6 – 18 based on our divergence analysis experiments. As showed in the Figure 8(a), even though the divergence value proves IDT as a prominent feature to distinguish spam from benign but in actual experiments, it also costs around 15% false positives to achieve 80% detection rate. This shows the inherent tradeoff between the accuracy and the efficiency of such behavioral divergence based detection.

4.2. Emails Per Recipient Based Detector

We have used a complete day as a bin to calculate the performance of EPR based detector. For a real time detection we can always shorten this bin size, but for our experiments we just wanted to understand how well this features works. The results are showed in Figure 9(a). The results show that the detection rate increases exponentially around threshold of 10 emails per recipient per day. The FPR ratio is high for this detector, i.e. 40%, if we want to detect with an accuracy of 90%.
4.3. Email Size Based Detector

In our divergence analysis we have come to know that the spammer tends to use similar template sizes for spam. Therefore, we are using variance in email sizes as a criteria to identify spam. We have used a bin size of 1 minute and calculate the variance of email sizes during that window. If the variance is high this means most likely it is benign but if the variance if low then it is classified as spam. The results in Figure 10(a) shows the ROC by varying the threshold in terms of percentage of variation on X-axis. These results coincide with the divergence results in Figure 4(a), that there is not enough leg room available to differentiate between spam and benign based on this feature alone. The maximum achievable accuracy was 80% but at the cost of high FPR, 50%.

5. Modeling Spam Economy

We are developing spam economic model to quantify spammer utility and constraining it with IDT to provide spammer cost analysis. We are using classical consumer theory of economics by considering spammer as consumer. Economics is the study of the choices people make about commodities (products or services) as a result of scarcity. In spam economy, the spammer is looking to choose a commodity that could maximize his/her utility. The process of developing an economic model has three components. 1) What is the commodity? 2) How does a consumer compare different commodities? 3) How to translate a commodity into utility?

5.1. Defining Commodity

The core variables which we focused on were the output rate and rental cost of the botnet, the cost of victim email addresses and the response...
rate. Subsets of the output rate made up of the botnet size, bot bandwidth amounts, and the spam message size, were areas where we found we had the most control along with parts of the response rate. The botnet size also varies with botnet growth and death rates of individual bots. We analytically model the world where the spammer lives and find that changing some of the variables slightly—such as the output rate—can significantly alter the spammer’s profit [25].

5.1.1. Output Rate

The primary variable that spammers and botmasters must deal with is the rate at which emails can be sent. This rate is actually made up of three key variables: a) the bandwidth of each bot in the botnet, b) the number of compromised computers in the botnet, and c) the size of each spam email.

**Botnet Size:** Botnets come in many sizes and they have varied over the years since the first botnet was discovered in 2003. The first botnets were very large with the number of bots well over one million. Botmasters soon discovered that large botnets were unmanageable, easily detected, and costly to maintain.

Today’s botnets attempt to remain hidden so their command and control servers are not shutdown, ISPs are not forced to end their contract, and vendors and security experts do not create patches to remove their controlling software. To remain stealthy, the size of 95% of existing botnets is now in the tens of thousands, with only a handful venturing into the hundreds of thousands. Shadowserver.org [26] constantly tracks the number of bots and botnets and they show that during the past 60 days the number of botnets has increased from 2750 to 3300. From the data that we have researched [27, 28], the average size of a botnet is around 20,000 bots with the median around 45,000 bots. Unfortunately, most bots are programmed to ‘sleep’ for long time periods. Bots wake up occasionally and ‘phone home’ for instructions. This constant on-off activity, in addition to the varying timezones of the global Internet, prevents researchers from gathering a completely accurate count of botnet sizes.

**Bot Bandwidth:** Spammer wants to use high bandwidth bots to send out more spam. As [29] states, the broadband transition to faster upload bandwidth via fibre could make the botnet problem much more severe. The potency of one infected computer on a fibre connection could be equivalent to 31 infected computers on DSL and 44 computers on cable networks. The current average upload speeds for broadband are 1 Mbit/s for DSL and 0.7
Figure 10: Figure (a) shows that ES can be a good differentiator if spammer uses large size templates by achieving decent DR high FPR. Our spamming cost metric can even be helpful in reducing FPR for less effective parameters as in Figure (b).

Mbit/s for cable, while fibre networks have an upload speed of 31 Mbit/s. An example is the Grum botnet which has a reported size of 50,000 bots and can output 2 billion spam emails per day. The Ozdok botnet, on the other hand, only has 35,000 bots, but can output 10 billion spam emails per day [30].

**Spam Message Size**: Our experimental results showed that the average email size sent by our trapped spam bots was 2,200 bytes. Based on these numbers, any future emails by spam sources in the same spam campaign could be predicted with a high degree of certainty to be around the same size. Obviously, the smaller the emails, more of them can be sent in a given period of time. Spammers have tried many different types of email formats in an attempt to get by spam filters. These include basic text, HTML and graphics.

### 5.1.2. Botnet Rental Cost and Duration

Botnets are usually not owned or operated by spammers. Most botnets are rented on a weekly basis, although some allow for a daily rental. There are currently over 500 botnets available for spammers to choose from, and payment is transferred through bank accounts or online payment services such as Paypal or eGold [29]. According to Spamhaus.org there are currently 111 *spam gangs* in operation around the world. These spam gangs are groups of one to five people, which makes the complete number of known high-volume spammers at around 300-400 people. Spamhaus.org tracks these spammers through their Registry of Known Spam Operators (ROKSO) list. The weekly rate for botnet rentals ranges from $50-$60 per 1,000 to 2,000 bots, which comes out to around 33 cents per bot [29].
5.1.3. Victim Email Addresses

In our experiment, we found that spammers do not repeat an email address more than \( x \) times during a spam campaign, where \( x \) is the number of email templates used. If they only have one email template, then an email address was not used more than once. This allowed the spammer to reach the largest number of people and at the same time avoid being blacklisted by a recipient’s mail server due to too much spam being sent to the same address. Email addresses are collected by programs that traverse the Internet recording any address they find on any page, especially blogs or other posting sites. These addresses are collected and sold to spammers at the rate of $100 per 10 million email addresses [29].

5.1.4. Response Rate

Spam would have disappeared years ago if people did not buy the products that the spammers are selling. According to Kanich et al. [2], the response rate from spam varies depending on the type of product being sold and the filters that are in place at the recipients mail server. Their research showed that after a 26 day spam campaign of pharmacy products, which included almost 350 million emails from Storm botnet servers, only 28 resulted in a sale which is a conversion rate of under 0.00001%. When they extrapolated their result to the rest of the Storm botnet they calculated the daily revenue for that campaign was roughly $7000.

5.2. Commodity Modeling

From a spammer’s perspective, maintenance cost is already a part of the per bot rent cost, so we can merge these two factors together. Similarly, the cost of the victim email addresses [29] is not a recurring cost, as the spammer can always reuse the previously acquired email addresses. One rule of thumb in defining a commodity is that the commodity must be recurring in nature, so we rule this cost out from commodity definition. We can merge the following factors, i.e., botnet size, bot bandwidth, spam mail size, into a derived factor, that we call the output rate. Let, \( \beta_i \) be the bandwidth of \( i^{th} \) bot, \( \delta \) be the average spam mail size, \( N \) be the number of available bots in the botnet and \( \lambda \) be the duration (quantity) of the bandwidth needed by the spammer, then we can formally define this output rate, \( c \), as follows:

\[
c = \sum_{i=1}^{N} \frac{\beta_i \cdot \lambda}{\delta}
\] (2)
where $c$ is the commodity of the model. The last factor, cost per bot, can be used to put a price tag on this commodity and this price tag is $p$ of the commodity $c$.

There can be infinite combinations of these factors and each represents a commodity. Let the collection of all these commodities within an economic system be $C$, where, $C = \{c_1, c_2, \ldots\}$ and the set of their price tags be, $P = \{p_1, p_2, \ldots\}$, where, $p_i$ is the price of the commodity $c_i$. A spammer can request multiple of these commodities, e.g., depending upon the different types of spam campaigns, a spammer may require different combinations of aggregated output rates. This means that for any spam activity, the spammer may acquire one commodity or a set of commodities. Each of this set of commodities is called a consumption bundle or a consumption set. Let a consumption bundle be $b$, e.g., $b_1 = \{c_1, c_2\}$ with an associated price $r_1 = \{p_1, p_2\}$, and set of all consumption bundles be $B$.

Like every consumer, the spammer also has a limited wealth, $W$ and he/she cannot afford to buy every possible consumption bundle. Then all those consumption bundles that a spammer can buy based on his/her wealth are called the competitive budget set, $B'$. This is formally defined as:

$$B' = \{\forall b_i \in B, b_i^T \times r_i \leq W\}. \tag{3}$$

where $b_i^T$ is simply the transpose of the consumption bundle $b_i$. In this equation $B'$ represents the affordable consumption bundles for the spammer. In the following section we discuss the choice structure of the spammer.

5.3. Defining Choice Behavior of The Spammer

The choice that a consumer makes in an economy is called the Preference Relation, $\succeq$. It is a binary relation on the set of consumption bundles alternatives available to the consumer. If, $b_1$ and $b_2$ are two consumption bundles, then $b_1 \succeq b_2$ means that $b_1$ is at least as good as $b_2$ and $b_1 \succ b_2$ means that $b_1$ is preferred over $b_2$. The role of the preference relation between consumption bundles is very critical as its absence makes an economic model unsolvable.

In our economic model, we are assuming a rational behavior of the spammer, which says, spammer will choose a consumption bundle that would yield higher utility. It is a fair assumption, but the commodity structure must adhere to certain properties before this assumption can be made. We discuss these properties and their justification for our commodity definition as follows.
1. Preference relation \( \succeq \) should be **rational**, i.e. it should be complete and transitive. **Completeness** property implies that \( \forall b_i, b_j \in B', (b_i \geq b_j) \lor (b_j \geq b_i) \). And, **Transitivity** property implies that \( \forall b_i, b_j, b_k \in B', ((b_i \geq b_j) \land (b_j \geq b_k)) \rightarrow (b_i \geq b_k) \). **Explanation:** Every commodity in our economic model can be quantifiable in the form of aggregate output rate, which is comparable among different commodities. Therefore, the commodity aggregate output rate conforms to this property.

2. Preference relation \( \succeq \) should be **monotone**. This property says, if two consumption bundles have the same aggregate rate then the consumption bundle with more commodities should be preferred, i.e. \( \forall b_i, b_j \in B'|b_i = b_2,(|b_i| > |b_j| \rightarrow b_i \geq b_j) \). **Explanation:** To defeat spammer, the research work in [12] proposed the concept of virtual bots (like honey pots) to lure malwares infecting fake bots. And in the end, a botnet would be consisted of virtual bots rather than the real bots, that could render the botnet useless. Now, based upon this defense mechanism, let’s assume that each commodity has a failure probability \( f \) and a commodity represents the usage of a botnet. Let this probability be the same across all commodities. Then, the failure probability of a consumption bundle will be the combination of all failure probabilities of each of its commodity. If \( b_1 = \{c_1, c_2\} \) and \( b_2 = \{c_1, c_3, c_4\} \), where, \( b_1 = b_2 \), then failure probabilities of \( b_1 \) and \( b_2 \) will be \( (f * f) \) and \( (f * f * f) \), respectively. This clearly indicates that the consumption bundle with more commodities is more reliable than the bundle with less number of commodities, if they have the same output rate. Therefore, this proves that our commodity definition also holds this monotone property.

3. The preference relation \( \succeq \) should be **convex** such as for every \( b_j \in B' \) the upper contour set is convex. It means, if \( b_i \succeq b_j \) and \( b_k \succeq b_j \), and \( b_i \neq b_k \), then \( \alpha b_i + (1 - \alpha) b_k \geq b_j \) for any \( \alpha \in [0, 1] \). **Explanation:** In standard economic theory, there are two reasons to impose this constraint: a) consumers typically like to consume mixed consumption bundles, like the monotone property, a bundle with more commodities is preferred; and b) It diminishes the marginal rate of substitution. Suppose, we have a function \( F(.) \) that calculates the utility of a consumption bundle. If there is a bundle \( b_1 = \{c_1, c_2\} \) then by diminishing the marginal rate of substitution, given by \( \frac{\partial F(b_1)}{\partial c_1} / \frac{\partial F(b_1)}{\partial c_2} \), the consumer requires more units of \( c_1 \) to remove one unit of \( c_2 \) to get the same utility. This gives stability to a consumption bundle by creating an area of indifference around it and indirectly gives a confidence to
the spammer in his/her selection. Same holds true in our commodity definition as the evasion due to diversity makes the associated cost of each substitution non-linear.

5.4. Defining Utility Function of The Spammer

There has to be a way to quantify or translate each selected commodity to a utility. According to the standard definition of utility:

**Definition 1.** A function $u(.)$ is a utility function representing preference relation $\succeq$, only if $\forall b_i, b_j \in B', b_i \succeq b_j \iff u(b_i) \geq u(b_j)$.

We use the properties discussed in the previous Section 5.3 i.e. the size of the consumption bundle and failure probability of a commodity, along with the commodity of the model to define a utility function as follows:

$$u(b_i) = \sum_{\forall c_j \in b_i} c_j + (|b_i| \ast (1 - f_j))$$

where $|b_i|$ gives the count of commodities in a consumption bundle and $f_j$ represents the failure probability of the commodity $c_j$. The objective of the spammer is to maximize his/her utility, that can be defined as:

$$\max u(b_i) \quad s.t. \quad b_i \in B' & b_i \neq \emptyset$$

6. Constraining Utility Function with Detection Features

The primary objective of the paper is to develop a sense of money into the spam detectors. Now, we will establish that missing link between the spam detection and the spam economy in this section. We will explain the changes in the economic model with respect to each of earlier mentioned detection features.

6.1. IDT Based Constraint

In order to maximize the use of commodity, the spammer would want to use 0 second IDT to send as many spam emails as possible. However, in order to avoid detection, the spammer would reside mixing this IDT, but the final number of spams can be calculated using mean value of IDT. Let $t$ be the IDT then $\mu(t)$ is the mean IDT. This increase in IDT reduces the
effective output rate of a commodity, which updates the basic commodity
definition in Equation 2 to the following as:

\[ \bar{c} = \sum_{i=1}^{N} \frac{\beta_i}{\delta + (\mu(idt) \ast \beta_i)} \]  \(6\)

The intuition of this equation is simple, the \(\beta_i\) represents the bits per second
(unit time) and \(\mu(idt)\) represents a pause, which can be translated into
the loss of bits that could have been sent otherwise. In other words hypotheti-
cally it can be taken as the increase in email size \(\delta\).

6.2. EPR Based Constraint

A spammer would want to send only a reasonable emails to any recipi-
ent to raise any suspicion. Let, \(\gamma\) be the total number of recipient’s email
addresses owned by the spammer. If a spammer can safely send only \(\kappa\)
number of emails to each recipient without causing any suspicion, then the total
number of emails that can be sent by the spammer will be \(\gamma \ast \kappa\). Therefore, ideally \(\gamma \ast \kappa > \max(b_i \in B')\). Even though the spammer may have
more wealth \((W)\), but EPR can potentially further constrain the competi-
tive budget set from Equation 2. The new constrained competitive budget
set is defined as follows:

\[ \tilde{B'} = \{ \forall b_i \in B, (b_i^T \times r_i \leq W) \land \left( \sum_{c_j \in b_i} c_j \leq (\gamma \ast \kappa \ast \theta) \right) \}. \]  \(7\)

6.3. ES Based Constraint

Our dataset analysis reveals only a handful of email templates used by
the spammer. The spammer must have to vary the size of the emails to stay
close to the benign behavior. But, the spammer cannot reduce the size of the
spam message because he/she cannot afford to leave out useful information.
So, a spammer would end up using a larger email size on average than the
one he/she anticipated to start with. This further restricts the effective
output rate in Equation 6. Let, the effective email size be \(\delta' \geq \delta\), then the
effective output rate will be change as:

\[ \bar{\omega} = \sum_{i=1}^{N} \frac{\beta_i}{\delta' + (\mu(idt) \ast \beta_i)} \]  \(8\)
7. Economic Metric Based Tuning

The utility function in Equation 4 provides the spammer utility in terms of the amount of spam emails that can be sent out. Currently, we are not using response rate and earning per response as part of the commodity because no such definite earning estimates are available to date. If such estimates are available we can easily extend this spammer’s utility function to reflect earnings as utility and this is one of our future tasks. In this paper we use this utility to calculate spammer’s cost in achieving this utility and then use our spam collection experiments statistics to calculate the botnet infrastructure required to yield this utility. We then use the IDT detection feature to limit spammer’s utility which is reflected into renting more infrastructure to achieve desired utility. This is how we calculate the increase in spammer’s cost as the detection rate increases.

From the botnet market [17, 10, 18], it costs around dollars 10 – 50 for a spammer to send one million spam emails. We are using $20 in our spammer’s cost calculations. We use our dataset experiment statistics to calculate per bot support required to achieve this one million figure. In our spam collection experiments, we have used 6 machines that collected almost 3 million spam emails in 60 days (two months), i.e., around 0.5 million emails are generated by each machine. This means that each machine has sent almost 8334 emails per day, then the average achieved IDT per machine per day is $\frac{60 \times 60 \times 24}{8334} = 10.32$ seconds. The average spam email size we have observed in our dataset is around 1000 bytes, this means that the approximate aggregate bandwidth used by each machine to send spam is almost 9 mega bytes per day ($1000 \times 8334$). If we assume that each bot will be offering same support then the spammer has to use 120 bots per day to send out one million spam emails. A spammer needs to send approximately 10 millions emails to get one positive response [2] therefore, we start with the spammer’s cost to acquire enough bots to send around 10 millions emails per day.

7.1. Improving IDT Based Detection Threshold

After constraining the spammer utility with IDT the effective utility of the spammer will be decreased, that requires spammer to spend more money to send out the desired 10 millions emails. This increase in cost with respect to increase in IDT value is plotted in the Figure 8(b). These results extend the results previously showed in Figure 8(a) with the spammer’s cost plotted on the secondary Y-axis. The line with circular dots (green line) represents the linear increase in spammer’s cost as the detection rate
increases. With this economic insight we do not need to tune our detector to achieve detection rate more than 80%, instead we can stop further below if we believe that the spammer’s cost is raised enough to defeat the objective of his/her spam campaign. We are working on to get those cost estimates that would render spam campaign useless for a spammer as another future task of this work.

7.2. Improving EPR Based Detection Threshold

According to the price of emails list in [17, 10], it cost $200 to $500 for spammer to acquire an email list of 1 Million recipients. In practical, the spammer can acquire millions of emails addresses but then he/she also has to send way more than 10 Million emails to get sufficient positive response to be profitable. In our experiment, we assume that a spammer has only 0.8 Million emails addresses to send 10 Million spam emails. The reason we are assuming these many email addresses because in our dataset analysis we have observed much less than 10 benign emails per day for any recipient. So, with 0.8 Million email addresses the spammer can achieve the target but it can also be restricted by the detector. We have calculated the cost on the secondary Y-axis using a very simple calculation. The value of the X-axis represents the allowed emails per recipient threshold, $\kappa$. Due to this restriction, if the spammer cannot send 10 Million emails from the rented infrastructure, this means the effective output rate of the network has reduced. This can be mapped to the situation if IDT has to be increased to avoid detection. So, by mapping reduced effective output rate to the associated IDT, we simply use the same technique from Section 7 to calculate the botnet cost. Again, the green line (with triangle) in Figure 9(b) shows the increase in cost for the spammer as the sensitivity of the detector is increasing. Again, if we are to stop at the threshold where the cost for the spammer is doubled, then we can stop at just 40% detection rate and at just 20% FPR, instead of whooping 40% FPR for 90% detection in Figure 9(b).

7.3. Improving ES Based Detection Threshold

The spammer tends to stay close to predefined template sizes to achieve estimated output rate from the network. But, to evade detection, the spammer has to introduce variance in email size. The spammer would not want to reduce the size as it might end-up not sending the required spam information to the user. Therefore, mostly the spammer will increase the size of the email to introduce variance. This will increase the average email size, thus, reduce the effective output rate of the spammer. But, in our evaluation we have seen that the spammer does not have to lose a lot due to
this constraint. Therefore, even in the secondary Y-axis of the Figure 10(b) the cost is not increasing a lot as in the case with previous detectors. But, still the cost indicator can help keeping the sensitivity of the detector at an affordable FPR rate. We would suggest an effective detection rate of 40% increase the cost of the spammer by 10%.

8. Conclusion and Future Work

The spam detectors suffer from the inherent tradeoff between accuracy and efficiency. Since, spam is all about earning money, in this paper we have introduced an economic metric by associating the detection accuracy of the detectors with the spammer’s cost. Firstly, we have developed a mechanism to identify effective spam detection features (like IDT) using $K$ directed divergence analysis, followed by the use of ROC curves to evaluate the performance of IDT based detector. Secondly, we have developed a spam economic model that calculates the spammer utility in terms of the number of spam emails sent by the spammer. Further constraining this utility with the IDT feature, we have added a cost sense into the detector to focus on increasing spammer’s cost rather than just the detection accuracy.

As a future task, we are planning to extend the utility function of the spammer to reflect his/her earnings as well by using both rational and irrational choice behaviors.


