Show Me the Money: Characterizing Spam-advertised Revenue

Chris Kanich* Nicholas Weaver† Damon McCoy* Tristan Halvorson*
Christian Kreibich† Kirill Levchenko*
Vern Paxson†‡ Geoffrey M. Voelker* Stefan Savage*

*Department of Computer Science and Engineering
University of California, San Diego

†International Computer Science Institute
Berkeley, CA

‡Computer Science Division
University of California, Berkeley

Abstract

Modern spam is ultimately driven by product sales: goods purchased by customers online. However, while this model is easy to state in the abstract, our understanding of the concrete business environment—how many orders, of what kind, from which customers, for how much—is poor at best. This situation is unsurprising since such sellers typically operate under questionable legal footing, with "ground truth" data rarely available to the public. However, absent quantifiable empirical data, "guesstimates" operate unchecked and can distort both policy making and our choice of appropriate interventions. In this paper, we describe two inference techniques for peering inside the business operations of spam-advertised enterprises: purchase pair and basket inference. Using these, we provide informed estimates on order volumes, product sales distribution, customer makeup and total revenues for a range of spam-advertised programs.

1 Introduction

A large number of Internet scams are "advertising-based"; that is, their goal is to convince potential customers to purchase a product or service, typically via some broad-based advertising medium.[1] In turn, this activity mobilizes and helps fund a broad array of technical capabilities, including botnet-based distribution, fast flux name service, and bulletproof hosting. However, while these same technical aspects enjoy a great deal of attention from the security community, there is considerably less information quantifying the underlying economic engine that drives this ecosystem. Absent grounded empirical data, it is challenging to reconcile revenue "estimates" that can range from $2M/day for one spam botnet [1], to analyses suggesting that spammers make little money at all [6]. This situation has the potential to distort policy and investment decisions that are otherwise driven by intuition rather than evidence.

In this paper we make two contributions to improving this state of affairs using measurement-based methods to estimate:

- *Order volume.* We describe a general technique—purchase pair—for estimating the number of orders received (and hence revenue) via on-line store order numbering. We use this approach to establish rough, but well-founded, monthly order volume estimates for many of the leading "affiliate programs" selling counterfeit pharmaceuticals and software.
- *Purchasing behavior.* We show how we can use third-party image hosting data to infer the contents of customer "baskets" and hence characterize purchasing behavior. We apply this technique to a leading spamvertized pharmaceutical program and identify both the nature of these purchases and their relation to the geographic distribution of the customer base.

In each case, our real contribution is less in the particular techniques—which an adversary could easily defeat should they seek to do so—but rather in the data that we used them to gather. In particular, we document that seven leading counterfeit pharmacies together have a total monthly order volume in excess of 82,000, while three counterfeit software stores process over 37,000 orders in the same time.

On the demand side, as expected, we find that most pharmaceuticals selected for purchase are in the "male-enhancement" category (primarily Viagra and other ED medications comprising 60 distinct items). However, such drugs constitute only 62% of the total, and we document that this demand distribution has quite a long tail; user shopping carts contain 289 distinct products, including surprising categories such as anti-cancer medications

[1] Unauthorized Internet advertising includes email spam, black hat search-engine optimization [26], blog spam [21], Twitter spam [4], forum spam, and comment spam. Hereafter we refer to these myriad advertising vectors simply as spam.

(Arimidex and Gleevec), anti-schizophrenia drugs (Seroquel), and asthma medications (Advair and Ventolin). We also discover significant differences in the purchasing habits of U.S. and non-U.S. customers.

Combining these measurements, we synthesize overall revenue estimates for each program, which can be well in excess of $1M per month for a single enterprise. To the best of our knowledge, ours is the first empirical data set of its kind, as well as the first to provide insight into the market size of the spam-advertised goods market and corresponding customer purchasing behavior.

We structure the remainder of this paper as follows. In § 2 we motivate the need for such research, explain the limitations of existing data, and provide background about how the spam-advertised business model works today. We discuss our *purchase pair* technique in § 3, validating our technique for internal consistency and then presenting order volume estimates across seven of the top pharmaceutical affiliate programs and three counterfeit software programs. We then explore the customer dynamics for one particular pharmaceutical program, EvaPharmacy, in § 4. We explain how to use image log data to identify customer purchases and then document how, where and when the EvaPharmacy customer base places its orders. We summarize our findings in § 5, devising estimates of revenue and comparing them with external validation. We conclude with a discussion about the implications of our findings in § 6.

2 Background

The security community is at once awash in the technical detail of new threats—the precise nature of a new vulnerability or the systematic analysis of a new botnet's command and control protocol—yet somewhat deficient in analyzing the economic processes that underlie these activities. In fairness, it is difficult to produce such analyses; there are innate operational complexities in acquiring such economic data and inherent uncertainties when reasoning about underground activities whose true scope is rarely visible directly.

However, absent a rigorous treatment, the resulting information vacuum is all too easily filled with opinion, which in turn can morph into "fact" over time. Though pervasive, this problem seemingly reached its zenith in the 2005 claim by US Treasury Department consultant Valerie McNiven that cybercrime revenue exceeded that of the drug trade (over $100 billion at the time) [11]. This claim was frequently repeated by members of the security industry, growing in size each year, ultimately reaching its peak in 2009 with written Congressional testimony by AT&T's chief security officer stating that cybercrime reaped "more than $1 trillion annually in illicit profits" [23]—a figure well in excess of the entire software industry and almost twice the GDP of Germany. Nay-sayers are similarly limited in their empirical evidence. Perhaps best known in this group are Herley and Florencio, who argue that a variety of cybercrimes are generally unprofitable. However, lacking empirical data, they are forced to use an economic meta-analysis to make their case [5, 6, 7].

Unfortunately, the answer to such questions matters. Without an "evidence basis", policy and investment decisions are easily distorted along influence lines, either over-reacting to small problems or under-appreciating the scope of grave ones.

2.1 Estimating spam revenue and demand

In this paper we examine only a small subset of such activity: spam-advertised counterfeit pharmacies and, to a lesser extent, counterfeit software stores. However, even here public estimates can vary widely. In 2005, one consultancy estimated that Russian spammers earned roughly US$2–3M per year [18]. However, in a 2008 interview, one IBM representative claimed that a single spamming botnet was earning close to $2M *per day* [1]. Our previous work studied the same botnet empirically, leading to an estimate of daily revenue of up to $9,500, extrapolating to $3.5M *per year* [10]. Most recently, a report by the Russian Association of Electronic Communication (RAEC) estimated that Russian spammers earned 3.7 billion rubles (roughly $125 million) in 2009 [12].

The demand side of this equation is even less well understood, relying almost entirely on opt-in phone or email polls. In 2004, the Business Software Alliance sponsored a Forrester Research poll to examine this question, finding that out of 6,000 respondents (spread evenly across the US, Canada, Germany, France, the UK and Brazil) 27% had purchased spam-advertised software and 13% had purchased spam-advertised pharmaceuticals [3]. If such data were taken at face value, the US market size for spam-advertised pharmaceuticals would exceed 30 million customers. Similar studies, one by Marshal in 2008 and the other sponsored by the Messaging Anti-Abuse Working Group (MAAWG) in 2009, estimate that 29% and 12%, respectively, of Internet users had purchased goods or services advertised in spam email [8, 19].

In our previous work on empirically quantifying revenue for such activities, our measurements were only able to capture a few percent of orders for sites advertised by a single botnet serving a single affiliate program, GlavMed [10]. Here, we aim to significantly extend our understanding, with our results covering *total order volume* for five of the six top pharmacy affiliate programs, and three of the top five counterfeit software affiliate programs. Moreover, to the best of our knowledge our analysis of EvaPharmacy is the first measurement-based ex-

amination of customer purchasing behavior, the demand component of the counterfeit pharmacy ecosystem.

2.2 How spam-advertised sites work

To provide context for the analysis in this paper, we first describe how modern spam is monetized and the ecosystem that supports it.

Today, spam of all kinds represents an outsourced marketing operation in service to an underlying sales activity. At the core are "affiliate programs" that provide retail content (e.g., storefront templates and site code) as well as back-end services (e.g., payment processing, fulfillment and customer support) to a set of client affiliates. Affiliates in turn are paid on a commission basis (typically 30–50% in the pharmaceutical market) for each sale they bring in via whatever advertising vector they are able to harness effectively. This dynamic is well described in Samosseiko's "Partnerka" paper [22] and also in our recent work studying the spam value chain [16].

Thus, while an affiliate has a responsibility to attract customers and host their shopping experience (which includes maintaining the contents of their "shopping cart"), once a customer decides to "check out" the affiliate hands the process over to the operators of the affiliate program.[2] Consequently, we would expect to find the order processing service shared across *all* affiliates of a particular program, regardless of the means used to attract customers. Indeed, as discussed below, our measurements of purchases from different members of the same affiliate confirm that the order numbers associated with the purchases come from a common pool. This finding is critical for our study because it means that side-effects in the order processing phase reflect the actions of *all sales activity* for an entire program, rather than just the sales of a single member.

On the back end, order processing consists of several steps: authorization, settlement, fulfillment, and customer service. Authorization is the process by which the merchant confirms, through the appropriate payment card association (e.g., Visa, MasterCard, American Express, Japan Credit Bureau, etc.), that the customer has sufficient funds. For the most common payment cards (Visa/MC), this process consists of contacting the customer's issuing bank, ensuring that the card is valid and the customer possesses sufficient funds, and placing a lien on the current credit balance. Once the good or service is ready for delivery, the merchant can then execute a settlement transaction that actualizes this lien, transferring money to the merchant's bank. Finally, fulfillment comprises packaging and delivery (e.g., shipping drugs directly from a foreign supplier or providing a Web site and password for downloading software). For our study, however, the key leverage lies in *customer service*. To support customer service, payment sites generate individual order numbers to share with the customer. In the next section, we describe how we can use the details of this process to infer the overall transaction rate, and ultimately revenue, of an entire affiliate program.

[2]This transfer typically takes the form of a redirection to a payment gateway site (with the affiliate's identity encoded in the request), although some sites also support a proxy mode so the customer can appear to remain at the same Web site.

3 Order volume

Underlying our *purchase pair* measurement approach is a model of how affiliate programs handle transactions, and, in particular, how they assign order numbers.

3.1 Basic idea

Upon placing an order, most affiliate programs provide a confirmation page that includes an "order number" (typically numeric, or at least having a clear numeric component) that uniquely specifies the customer's transaction. For purchases where an order number does not appear on the confirmation page, the seller can provide one in a confirmation email (the common case), or make one available via login to the seller's Web site. The order number allows the customer to specify the particular purchase in any subsequent emails, when using customer support Web sites, or when contacting online support via email, IM or live Web chat. For the purchases we made, we found that the seller generally provides the order number *before* the authorization step (indeed, even before merchant-side fraud checks such as Address Verification Service), although purely local checks such as Luhn digit validation are frequently performed first. Accordingly, we can consider the creation of an order number only as evidence that a customer *attempted* an order, not that it successfully concluded. Thus, the estimates we form in this work reflect an *upper bound* on the transaction rate, including transactions declined during authorization or settlement.[3]

The most important property for such order numbers is their *uniqueness*; that each customer order is assigned a singular number that is distinguished over time without the possibility of aliasing. While there are a vast number of ways such uniqueness could be implemented (e.g., a pseudo-random permutation function), the easiest approach by far is to simply increment a global variable for each new order. Indeed, the serendipitous observation that motivated our study was that multiple purchases made from the same affiliate program produced

[3]In 2008, Visa documented that card-not-present transactions such as e-commerce had an issuer decline rate of 14% system-wide [25]. In addition, it seems likely that some orders are declined at the merchant's processor due to purely local fraud checks (such as per-card or per-address velocity checks or disparities between IP address geolocation versus shipping address).

order numbers that appeared to *monotonically increase* over time. Observing the monotonic nature of this sequence, we hypothesized that order number allocation is implemented by serializing access to a single global variable that is incremented each time an order is made; we call this the *sequential update hypothesis*. To assess this hypothesis, we examined source code for over a dozen common e-commerce platforms (e.g., Magento, X-cart, Ubercart, and Zen-cart [17, 24, 27, 28]), finding ubiquitous use of such a counter, typically using an SQL autoupdate field, but sometimes embodied explicitly in code.

Given use of such a global sequential counter, the *difference* between the numbers associated with orders placed at two points in time reflects the total number of orders placed during the intervening time period. Thus, from any *pair* of purchases we can extract a measurement of the total transaction volume for the interval of time between them, even though we cannot directly witness those intervening transactions. Figure 1 illustrates the methodology using a concrete example. This observation is similar in flavor to the analysis used in blind/idle port scanning (there the sequential increment of the IP identification field allows inference of the presence of intervening transmissions) [2]. It then appears plausible that this same purchase-pair approach might work across a broad range of spam-advertised programs, a possibility that we explore more thoroughly next.

3.2 Data collection

To evaluate this approach requires that we first identify which sites advertise which affiliate programs, and then place repeated purchases from each. We describe how we gathered each of these data sets in this section.

Program data

In prior work, we developed a URL crawler to follow the embedded links contained in real-time feeds of email spam (provided by a broad range of third-party antispam partners) [16]. The crawler traverses any redirection pages and then fetches and renders the resulting page in a live browser. We further developed a set of "page classifiers" that identify the type of good being advertised by analyzing the site content, and, in most cases, the particular affiliate program being promoted. We developed specific classifiers for over 20 of the top pharmaceutical programs (comprising virtually all sites advertised in pharmaceutical spam), along with the four most aggressively spam-advertised counterfeit software programs.

After placing multiple test orders with nine of these pharmaceutical programs, we identified seven with strictly incrementing order numbers.[4] Five of these (Rx–Promotion, Pharmacy Express (aka Mailien), GlavMed, Online Pharmacy and EvaPharmacy) together constituted two-thirds of all sites advertised in the roughly 350 million distinct pharmaceutical spam URLs we observed over three months in late 2010. We found the sixth, 33drugs (aka DrugRevenue), and seventh, 4RX, less prevalent in email spam URLs, but they appear to be well advertised via search engine optimization (SEO) techniques [15]. We did a similar analysis of counterfeit software programs, finding three (Royal Software, EuroSoft, and SoftSales) with the appropriate order-number signature. While counterfeit software is less prevalent in total spam volume, these three programs constitute over 97% of such sites advertised to our spam collection apparatus during the same 3-month period. For the remainder of this paper we focus exclusively on these ten programs, although it appears plausible that the same technique will prove applicable to many smaller programs, and also to programs in other such markets (e.g., gambling, fake antivirus, adult).

Order data

We collected order data in two manners: actively via our own purchases and opportunistically, based on the purchases of others. First and foremost are our own purchases, which we conducted in two phases. The first phase arose during a previous study, during which we executed a small number of test purchases from numerous affiliate programs in January and November of 2010 using retail Visa gift cards. Of these, 46 targeted the ten programs under study in this paper. The second phase (comprising the bulk of our active measurements) reflects a regimen of purchases made over three weeks in January and February 2011 focused specifically on the ten programs we identified above.

When placing these orders, we used multiple distinct URLs leading to each program (as identified by our page classifiers). The goal of this procedure was to maximize the likelihood of using distinct affiliates to place purchases in order to provide an opportunity to determine whether different affiliates of a given program make use of different order-processing services.

Successfully placing orders had its own set of operational challenges [9]. Except where noted, we performed all of our purchases using prepaid Visa credit cards provided to us in partnership with a specialty issuer, and funded to cover the full amount of each transaction. We used a distinct card for each purchase and went to considerable lengths to emulate real customers. We used valid names and associated residential shipping addresses, placed orders from a range of geographically

[4] Of the two programs that we did not select, ZedCash used several different strictly increasing order number subspaces that would complicate our analysis and decrease accuracy, while World Pharmacy order numbers appeared to be the concatenation of a small value with the current Unix timestamp, which would thwart our analysis altogether.

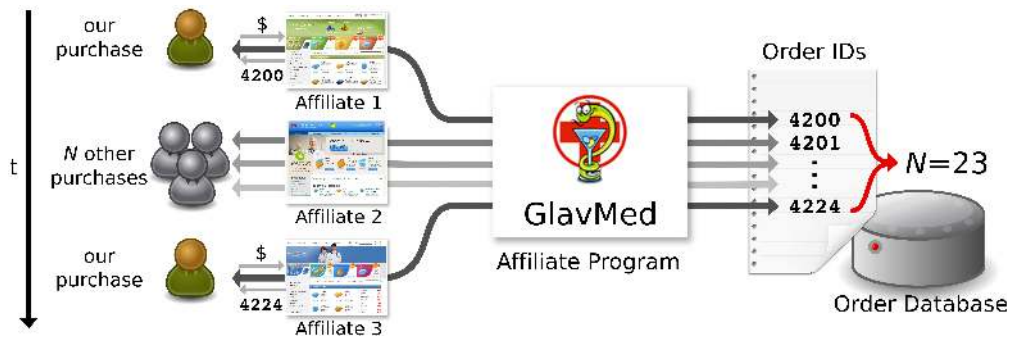

Figure 1: How the purchase pair technique works. In this hypothetical situation, two measurement purchases are made that bracket some number of intervening purchases made by real customers. Because order number allocation is implemented by a serialized sequential increment, the difference in the order numbers between measurement purchases, $N = 23$, corresponds to the total number of orders processed by the affiliate program in the intervening time.

proximate IP addresses, and provided a unique email address for each order. We used five contact phone numbers for order confirmation, three from Google Voice and two via prepaid cell phones, with all inbound calls routed to the prepaid cell phones. In a few instances we found it necessary to place orders from IP addresses closely geolocated to the vicinity of the billing address for a given card, as the fraud check process for one affiliate program (EuroSoft) was sensitive to this feature. Another program (Royal Software) would only accept one order per IP address, requiring IP address diversity as well.

In total we placed 156 such orders. We scheduled them both periodically over a three-week period as well as in patterns designed to help elucidate more detail about transaction volume and to test for internal consistency, as discussed below.

Finally, in addition to the raw data from our own purchase records, we were able to capture several purchase order numbers via forum scraping. This opportunity arose because affiliate programs typically sponsor online forums that establish a community among their affiliates and provide a channel for distributing operational information (e.g., changes in software or name servers), sharing experiences (e.g., which registrars will tolerate domains used to host pharmaceutical stores), and to raise complaints or questions. One forum in particular, for the GlavMed program, included an extended "complaint" thread in which individual affiliates complained about orders that had not yet cleared payment processing (important to them since affiliates are only paid for each settled transaction that they deliver). These affiliates chose to document their complaints by listing the order number they were waiting for, which we determined was in precisely the same format and numeric range as the order numbers presented to purchasers. By mining this forum we obtained 122 numbers for past orders, including orders dating back to 2008.

Affiliate Program	Phase 1 (1/10 – 11/10)	Phase 2 (1/11 – 2/11)
Rx–Promotion	7	27
Pharmacy Express	3	9
GlavMed	12	14
Online Pharmacy	5	16
EvaPharmacy	7	16
33drugs	4	16
4RX	1	13
EuroSoft	3	25
Royal Software	2	9
SoftSales	2	11

Table 1: Active orders placed to sites of each affiliate program in the two different time phases of our study. In addition, we opportunistically gathered 122 orders for GlavMed covering the period between 2/08 and 1/11.

Note that this data contains an innate time bias since the date of complaint inevitably came a while later than the time of purchase (unlike our own purchases). For this reason, we identify opportunistically gathered points distinctly when analyzing the data. We will see below that the bias proves to be relatively minor.

We summarize the total data set in Table 1. It includes order numbers from 202 active purchases and 122 opportunistically gathered data points.

3.3 Consistency

While our initial observations of monotonicity are quite suggestive, we need to consider other possible explanations and confounding factors as well. Here we evaluate the data for *internal consistency*—the degree to which the data appears best explained by the *sequential update hypothesis* rather than other plausible explanations. At the end of the paper we also consider the issue of *external consistency* using "ground truth" revenue data for one program.

Sequential update

The fundamental premise underlying our purchase-pair technique is that order numbers increment sequentially for each attempted order. The monotone sequences that we observe accord with this hypothesis, but could arise from other mechanisms. Alternate interpretations include that updates are monotone but not sequential (e.g., incrementing the order number by a small, varying number for each order) or that order numbers are derived from timestamps (i.e., that each order number is just a normalized representation of the time of purchase, and does not reflect the number of distinct purchase attempts).

To test these hypotheses, we executed back-to-back orders (i.e., within 5–10 seconds of one another) for each of the programs under study. We performed this measurement at least twice for all programs (excepting EvaPharmacy, which temporarily stopped operation during our study). For eight of the programs, every measurement pair produced a sequential increment. The GlavMed program also produced sequential increments, but we observed one measurement for which the order number incremented by two, likely simply due to an intervening order out of our control. Finally, we observed no sequential updates for Rx–Promotion even with repeated back-to-back purchase attempts. However, upon further examination of 35 purchases, we noticed that order numbers for this program are always odd; for whatever reason, the Rx–Promotion order processing system increments the order number *by two* for each order attempt. Adjusting for this deviation, our experiments find that on finer time scales, every affiliate program behaves consistently with the sequential update hypothesis.

We need however to consider an alternate hypothesis for this same behavior: that order numbers reflect normalized representations of timestamps, with each order implicitly serialized by the time at which it is received. This "clock" model does not appear plausible for fine-grained time scales. Our purchases made several seconds apart received sequential order numbers, which would require use of a clock that advances at a somewhat peculiar rate—slowly enough to risk separate orders receiving the same number and violating the uniqueness property.

A possible refinement to the clock model would be for a program to periodically allocate a block of order numbers to be used for the next T seconds (e.g., for $T = 3{,}600$), and after that time period elapses, advancing to the next available block. The use of such a hybrid approach would enable us to analyze purchasing activity over fine-grained time scales. But it would also tend towards misleading over-inflation of such activity on larger time scales, since we would be comparing values generated across gaps.

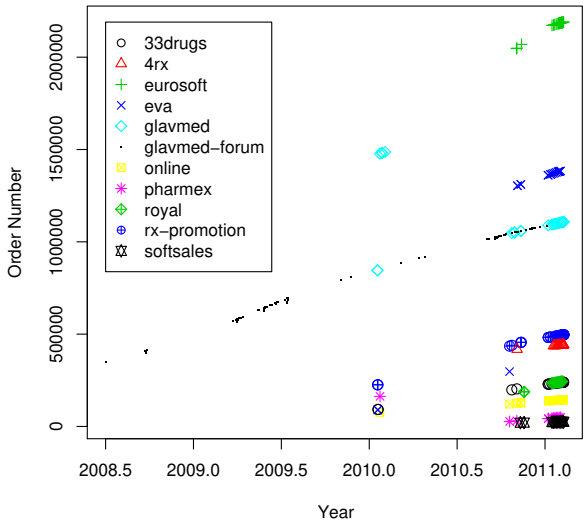

Figure 2: Order numbers (y-axis) associated with each affiliate program versus the time of attempted purchase (x-axis).

We test for whether the order numbers in our data fit with a clock model as follows. First, we consider the large-scale behavior of order numbers as seen across the different affiliate programs. Figure 2 plots for each program the order number associated with a purchase attempt made at a given time. We plot each of the 10 affiliate programs with a separate symbol (and varying shades, though we reuse a few for programs whose numbers are far apart). In addition, we plot with black points the order numbers revealed in the GlavMed discussion forum.

Three basic points stand out from the plot. First, all of the programs use order numbers distinct from the others. (We verified that neither of those closest together, 33drugs and Royal Software, nor Pharmacy Express and SoftSales, overlap.) Thus, it is not the case that separate affiliate programs share unified order processing.

Second, the programs nearly always exhibit monotonicity even across large time scales, ruling out the possibility that some programs occasionally reset their counters. (We discuss the outliers that manifest in the plot below.)

Third, the GlavMed forum data is consistent with our own active purchases from GlavMed. In addition, the data for both has a clear downward concavity starting in 2009—inconsistent with use of clock-driven batches, but consistent with the sequential update hypothesis. Assuming that the data indeed reflects purchase activity, the downward concavity also indicates that the program has been losing customers, a finding consistent with mainstream news stories [13].

We lack such extensive data for the other programs, but can still assess their possible agreement with use

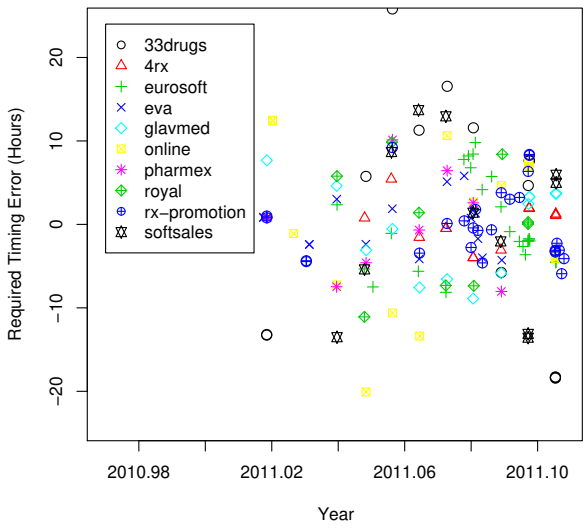

Figure 3: The amount of error—either in our measurement process, or due to batching of order numbers—required for each measurement in 2011 to be consistent with the Null Hypothesis that order numbers are derived from a clock that advances at some steady rate. Note that the y-axis is truncated at ± 24 hrs, though additional points lie outside this range.

of clock-driven batches, as follows. For each program, we consider the purchases made in 2011. We construct a least-squares linear fit between the order numbers of the purchases and the time at which we made them. If the order numbers come from clock-driven batches (the Null Hypothesis), then we would expect that all of the points associated with our purchases to fall near the fitted line. Accordingly, for each point we compute how far we would have to move it along the x-axis so that it would coincide with the line for its program. If the Null Hypothesis is true, then this deviation in time reflects the *error* that must have arisen during our purchase measurement: either due to poor accuracy in our own time-keeping, or because of the granularity of the batches used by the program for generating order numbers.

Figure 3 plots this residual error for each affiliate program. For example, in the lower right we see a point for a 33drugs purchase made in early February 2011. If the Null Hypothesis holds, then the purchaser's order number reflects a value that should have appeared 18 hours earlier than when we observed it. That is, either we introduced an error of about 18 hours in recording the time of that purchase; or the program uses a batch-size of 18+ hours; or the Null Hypothesis fails to hold.

For all ten of the affiliate programs, we find many purchases that require timing errors of many hours to maintain consistency with the Null Hypothesis. (Note that we restrict the y-axis to the range ± 24 hr for legibility, although we find numerous points falling outside that range as well.) In addition, we do not discern any temporal patterns in the required errors, such as would be the case if the least-squares fit was perturbed by an outlier. Finally, if we extend the analysis out to November 2010 (not shown), we find that the required error *grows*, sometimes to 100s of hours, indicating that the discrepancy does not result from a large batch size such as $T = 1$ day.

Given this evidence, we reject the Null Hypothesis that the order numbers derive from a clock-driven mechanism. We do however find the data consistent with the sequential update hypothesis, and so proceed from this point on the presumption that indeed the order numbers grow sequentially with each new purchase attempt.

Payment independence

We placed most of our orders using cards underwritten by Visa. We selected Visa because it is the dominant payment method used by these affiliate programs (few accept MasterCard, and fewer still process American Express). However, it is conceivable that programs allocate distinct order number ranges for each distinct type of payment. If so, then our Visa-based orders would only witness a subset of the order numbers, leading us to underestimate the total volume of purchase transactions. To test this question, we acquired several prepaid MasterCard cards and placed orders at those programs that accept MasterCard (doing so excludes Rx–Promotion, GlavMed, 4RX and Online Pharmacy). In each case, we found that Visa purchases made directly before and after a MasterCard purchase produced order numbers that precisely bracketed the MasterCard order numbers as well.

Outliers

Out of the 324 samples in our dataset, we found a small number of outliers (six) that we discuss here. Almost all come from the GlavMed program. The outliers fall into two categories: two singleton outliers completely outside the normal order number range for the program, and one group of four internally consistent order numbers that were slightly outside the expected range, violating monotonicity. We discuss these in more detail here, as well as their possible explanations.

The first singleton outlier was a purchase placed at a Web site that is clearly based on the SE2 engine built by GlavMed. However, the returned order number was close to 16000 when co-temporal orders from all other GlavMed sites returned orders closer to 1080000. The site differs in a number of key features, including a unique template not distributed in the standard package made available to GlavMed affiliates, a different support phone number, different product pricing, and purchases processed via a different acquiring bank than used by all other GlavMed purchases. Taken together, we believe

this reflects a site that is simply using the SE2 engine, but is not in fact associated with the GlavMed operation.[5]

The second outlier occurred in a very early (January 2010) purchase from a Pharmacy Express affiliate, which returned an order number much higher than any seen in later purchases. We have no clear explanation for this incongruity, and other key structural and payment features match, but we note that the order numbers returned in all subsequent Pharmacy Express transactions are only five digits long, and that over nine months pass between this initial outlier and all subsequent purchases. Consequently, we might reasonably explain the discrepancy by a decision to reset the order number space at some point between January and October.

Finally, we find a group of four early GlavMed purchases whose order numbers are roughly the same magnitude, but occur out of sequence (i.e., given the rate of growth seen in the other GlavMed order numbers, these four are from a batch that will only be used sometime in 2013). These all occurred together in the last two weeks of January 2010. This small outlier group remains a mystery, and suggests either that GlavMed might maintain a parallel order space for some affiliates, or that they reflect a "counterfeit" GlavMed operation. The remaining 21 GlavMed purchase samples, as well as the 122 opportunistically gathered order numbers (occurring both before and after January 2010), all use consistent order numbering.

While we cannot completely explain these few outliers, they represent less than 2% percent of our dataset. We also have found no unexplained instances within the last 12 months. We remove these six data points in the remainder of our analysis.

3.4 Order rates

Under these assumptions, we can now estimate the rate of orders seen by each enterprise. Figure 4 plots the 2011 data points for each of the 10 programs. We also plot the least squares linear interpolation as well as the slope parameter of this line—corresponding to the number of orders received per day on average. During this time period, daily order rates for pharmacy programs vary from a low of 227 for Rx–Promotion (recall that their order IDs increment by two for each order) up to a high of 887 for EvaPharmacy (software programs range between 49 and 749). Together, these reflect a monthly volume of over 82,000 pharmaceutical orders and over 37,000 software orders. Again, these numbers reflect upper bounds on completed orders, since undoubtedly some fraction of these attempted orders are declined; however, it seems clear that order volume is substantial.

We also note that while order volume is quite consistent across January and February, there are significant fall offs for some programs when compared to the data gathered earlier. For example, during 2010, the average number of Rx–Promotion orders per day was 385, 70% greater than during the first two months of 2011. Similarly, 2011 GlavMed orders are off roughly 20% from their 2010 pace, and EvaPharmacy saw a similar decline as compared to October and November of that year. Other programs changed little and maintained a stable level of activity.

4 Purchasing behavior

While the previous analysis demonstrates that pharmaceutical affiliate programs are receiving a significant volume of orders, it reveals little about the source of these orders or their contents. In this section, we use an opportunistic analysis of found server log data to explore these issues for one such affiliate program.

4.1 EvaPharmacy image hosting

In particular, we examine EvaPharmacy, a "top 5" spam-advertised pharmacy affiliate program.[6] In monitoring EvaPharmacy sites we observed that roughly two thirds "outsourced" image hosting to compromised third-party servers (typically functioning Linux-based Web servers). This behavior was readily identifiable because visits to such sites produced HTML code in which each image load was redirected to another server—addressed via raw IP address—at port 8080.

We contacted the victim of one such infection and they were able to share IDS log data in support of this study. In particular, our dataset includes a log of HTTP request streams for a compromised image hosting server that was widely used by EvaPharmacy sites over five days in August of 2010. While the raw IP addresses in our dataset have been anonymized (consistently), they have first been geolocated (using MaxMind) and these geographic coordinates are available to us. Thus, we have city-level source identifiability as well as the contents of HTTP logs (including timestamp, object requested, and referrer).

Through repeated experimentation with live EvaPharmacy sites, we inferred that the site "engine" can use dynamic HTML rewriting (similar to Akamai) to rewrite embedded image links on a *per visit* basis. On a new visit (tracked via a cookie), the server selects a set of five compromised hosts and assigns these (apparently in a quasi-random fashion) to each embedded image link served. During the five-day period covering our log data, our crawler observed 31 distinct image servers in use.

[5] We have found third parties contracting for custom GlavMed templates on popular "freelancer" sites, giving reason to believe that independent innovation exists around the SE2 engine created by GlavMed.

[6] Our page classifiers [16] identified EvaPharmacy in over 8% of pharmacy sites found in spam-advertised URLs over three months, with affiliates driving traffic to over 11,000 distinct domains.

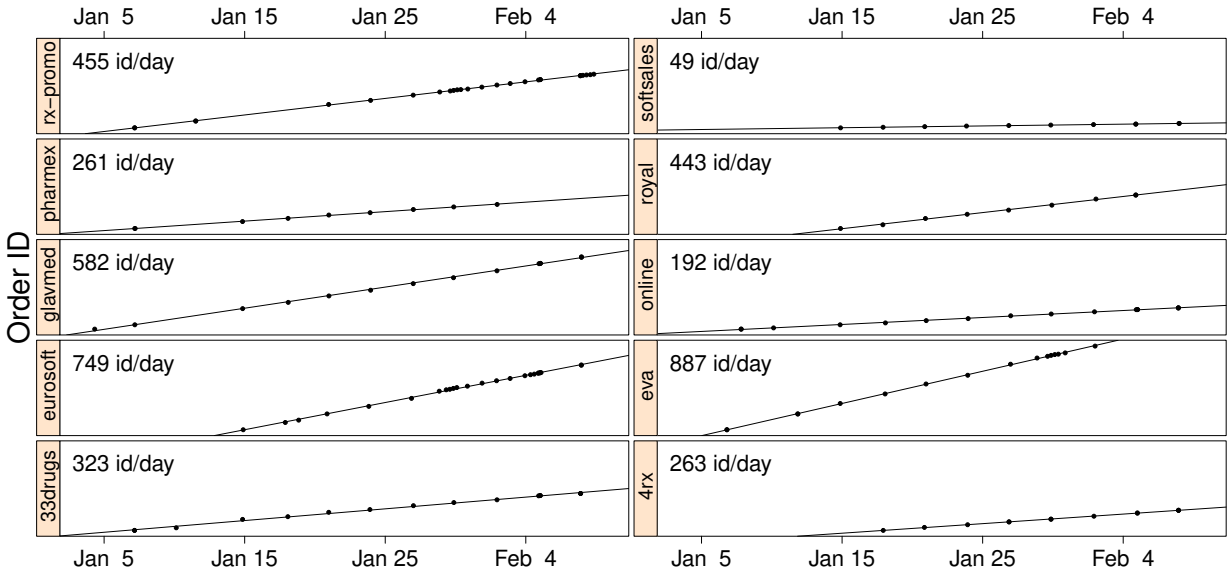

Figure 4: Collected data points and best fit slope showing the inferred order rate for ten different spam-advertised affiliate programs. Order numbers are zero-normalized and the vertical scale of each plot is identical.

However, our particular server was apparently disproportionately popular, as it appears in 31% of all contemporaneous visits made by our URL crawler (perhaps due to its particularly good connectivity). In turn, each image server hosts an nginx Web proxy able to serve the entirety of the image corpus.

4.2 Basket inference

Since the log we use is limited to embedded Web page images, and in fact only includes one fifth of the images fetched during a particular visit, there are considerable challenges involved in inferring item selection purely from this data. We next discuss how this inference technique works (illustrated at a high level in Figure 5) as well as its fundamental limitations.[7]

We mapped out the purchasing workflow involved in ordering from an EvaPharmacy site, and observed that all purchases involve visiting four key kinds of pages in order: landing, product, shopping cart, and checkout. The landing page generally includes over 40 distinct embedded images. Thus, even though images are split among five servers, it is highly likely that multiple objects from each landing page are fetched via our server (each with a referrer field identifying the landing page from which it was requested).[8] We observe 752,000 distinct IP addresses that visited and included referrer information during our five-day period.

When a visitor selects a particular drug from the landing page, the reply takes them to an associated product page. This page in turn prompts them to select the particular dosage and quantity they wish to purchase. The precise construction of product pages differs between the set of site templates (i.e., storefront brands) used by EvaPharmacy. However, all include at least a few new images not found on the landing page, and the most popular template fetches five additional images. The number of additional images varies on a per-template basis, not a per-product basis within each template. Thus, for some templates we may have less opportunity to observe what product the user selects, but this does not affect our estimate of the *distribution* of products selected, because the diminished opportunity is not correlated with particular products.

Next, upon selecting a product, the user is taken to the shopping cart page, which again includes a large number (often a dozen or more) of new images representing *product recommendations*. We observe 4,879 cart visits from 3,872 distinct IP addresses. This allows us to estimate a product-selection conversion rate: the fraction of visitors who select an item for purchase. Based on the total number of visitors where we have referrer information, the conversion percentage on an IP basis is 0.5%.[9] Of these, 3,089 cart additions have preceding visits to prod-

[7] This general approach is similar in character to Moore and Clayton's inference of phishing page visits from Webalizer logs [20].

[8] We validated this observation using our crawled data, which showed that the landing pages using :8080 image hosting always used five distinct servers. Thus, any image server assigned to a particular visit is guaranteed to see the landing page load for that visit.

[9] For comparison, in our previous work we measured a visit-to-product-selection conversion rate of 2% [10].

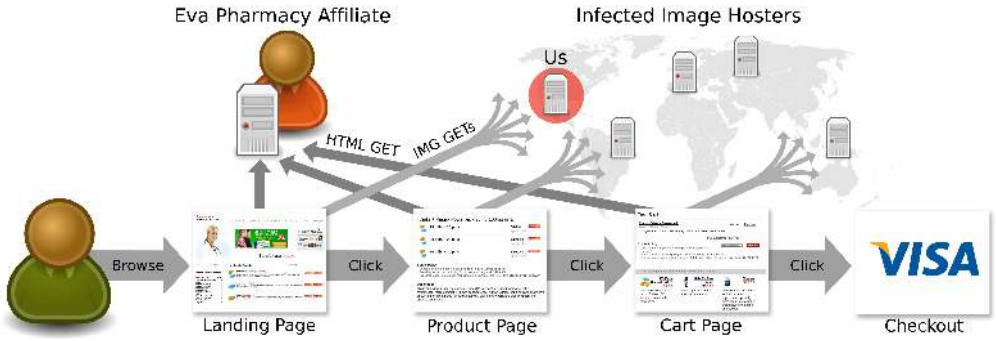

Figure 5: How a user interacts with an EvaPharmacy Web site, beginning with the landing page and then proceeding to a product page and the shopping cart. The main Web site contains embedded images hosted on separate compromised systems. When a browser visits such pages, the referrer information is sent to the image hosting servers for every new image visited.

uct pages, which allows us to infer the selected product. To quantify overall shopping cart addition activity, we compare the total number of visits to the number of visits to the shopping cart page. To quantify individual item popularity, we examine the subset of visits for which the customer workflow allows us to infer which specific item was added to the cart.

There are three key limitations to this approach. First and foremost, the final page in the purchasing workflow—the checkout page—generally does not include unique image content, and thus does not appear in our logs (even if it did, our approach could not determine whether checkout completed correctly). Thus, we can only observe that a user inserted an item into their cart, but not that they completed a purchase attempt. In general, this is only an issue to the degree that shopping cart abandonment correlates with variables of interest (e.g., drug choice). The second limitation is that pages typically use the same image for all dosages and quantities on a given product page, and therefore we cannot distinguish these features (e.g., we cannot distinguish between a user selecting 120 tablets of 25mg Viagra tablets vs. an order of 10 tablets, each of 100mg). Finally, we cannot disambiguate multiple items selected for purchase. When a user visits a product page followed by the shopping cart page, we can infer that they selected the associated product. However, if the visitor then *continues* shopping and visits additional product pages, we cannot determine whether they added these products or simply examined them (subsequent visits to the shopping cart page add few new recommended products; recommendations appear based on the first item in the cart). We choose the conservative approach and only consider the products that we are confident the user selected, which will cause us to under-represent those drugs typically purchased together.

Another issue is that pharmacy formularies, while largely similar, are not identical between programs. In particular, some pharmacy programs (e.g., Online Pharmacy) offer Schedule II drugs (e.g., Oxycodone and Vicodin). However, since EvaPharmacy does not sell such drugs, our data does not capture this category of demand.

Finally, our dataset also has potential bias due to the particular means used to drive traffic to it. We found that 45 of the 50 top landing pages observed in the hosting data also appeared in our spam-driven crawler data, demonstrating directly that these landing pages were advertised through email spam. While these pages could also be advertised using less risky methods such as SEO, this seems unlikely since spam-advertised URLs are swiftly blacklisted [14]. Thus, we suspect (but cannot prove) that our data may *only* capture the purchasing behavior for the spam-advertised pharmacies; different advertising vectors could conceivably attract different demographics with different purchasing patterns.

Given these limitations, we now report the results of two analyses: product popularity (what customers buy) and customer distribution (where the money comes from).

4.3 Product popularity

Our first analysis focuses on simple popularity: what individual items users put into their shopping carts (Table 3a) and what broad (seller-defined) categories of pharmaceuticals were popular (Table 3b) during our measurement period. Although naturally dominated by the various ED and sexually-related pharmaceuticals, we find a surprisingly long tail; indeed, 38% of all items added to the cart were not in this category. We observed 289 distinct products, including popular mass-market products such as Zithromax (31), Acomplia (27), Nexium (26), and Propecia (27); but also Cipro (11; a commonly prescribed antibiotic), Actos (6; a treatment for Type 2 diabetes), Buspar (12; anti-anxiety), Seoquel (9; anti-schitzophrenia), Clomid (8; ovulation inducer), and Gleevec (1; used to treat Leukemia and other cancers).

Figure 6: The geographic distribution of those who added an item to their shopping cart.

Country	Visits	Cart Additions	Added Product
United States	517,793	3,707	0.72%
Canada	50,234	218	0.43%
Philippines	42,441	39	0.09%
United Kingdom	39,087	131	0.34%
Spain	26,968	59	0.22%
Malaysia	26,661	31	0.12%
France	18,541	37	0.20%
Germany	15,726	56	0.36%
Australia	15,101	86	0.57%
India	10,835	17	0.16%
China	8,924	30	0.34%
Netherlands	8,363	21	0.25%
Saudi Arabia	8,266	36	0.44%
Mexico	7,775	17	0.22%
Singapore	7,586	17	0.22%

Table 2: The top 15 countries and the percentage of visitors who added an item to their shopping cart.

This in turn explains why such online pharmacies maintain a comprehensive inventory: not only does a full formulary lend legitimacy, but it also represents a significant source of potential revenue.

We also comprehensively crawled an EvaPharmacy site for pricing data and calculated the *minimum* estimated revenue per purchase (also shown for the top 18 products in Table 3a). Combining this data with our measurement of item popularity, we calculate a minimum weighted-average item cost of $76 plus $15 for shipping and handling. This weighted average assumes visitors always select the minimum-priced item for any given purchase, and that the final purchases have the same distribution as for items added to the user's shopping cart.

4.4 Customer distribution

We next examine the geographic component of the EvaPharmacy customer base. Figure 6 shows the geolocated origin for all shopping cart additions. We observe that EvaPharmacy has a vast advertising reach, producing site visits from 229 distinct countries or territories. However, this reach is not necessarily all that useful: the population *actively engaging* with EvaPharmacy sites and placing orders is considerably less diverse than the superset simply visiting (perhaps inadvertently or due to curiosity). For example, the Philippines constitutes 4% of the visitors, but only 1% of the additions to the shopping cart. Overall, countries other than the U.S., Canada, and Western Europe generate 29% of the visitors but only 13% of the items added to the shopping cart. Conversely, the vast majority of shopping cart insertions originate from the U.S. and Canada (80%) or Europe (6%), reinforcing the widely held belief that spam-advertised pharmaceuticals are ultimately funded with Western Dollars and Euros.

The United States dominates both visits (54%) and cart additions (76%), and moreover has the highest rate of conversion between visit and shopping cart insertion (0.72%). Table 2 well illustrates this, listing the activity from the countries originating the most visits. This observation reinforces the conclusion that non-Western audiences offer ineffective targets for such advertising.

Finally, we also notice significant differences between the drug selection habits of Americans compared to customers from Canada and Western Europe. In particular, we divide the EvaPharmacy formulary into two broad categories: lifestyle drugs (defined as drugs commonly used recreationally, including "male-enhancement" items plus Human Growth Hormone, Soma and Tramadol) and non-lifestyle (all others, including birth control pills). We find that while U.S. customers select non-lifestyle items 33% of the time, Canadian and Western-European customer selections concentrate far more in the lifestyle category—only 8% of all items placed in a shopping cart are non-lifestyle items. We surmise that this discrepancy may arise due to differences in health care regimes; drugs easily justified to a physician may be fully covered under state health plans in Canada and Western Europe, leaving an external market only for lifestyle products. Conversely, a subset of uninsured or under-insured customers in the U.S. may view spam-advertised, no-prescription-required pharmacies as a competitive market for meeting their medical needs. To further underscore this point, we observe that 85% of *all* non-lifestyle drugs are selected by U.S. visitors.

5 Revenue estimation

Combining the results from estimates on the order rate per program and estimates of the shopping cart makeup, we now estimate total revenue on a per-program basis.

5.1 Average price per order

The revenue model underlying our analysis is simple: we multiply the estimated order rate by the average price per order to arrive at a total revenue figure over a given unit

Product	Quantity	Min order
Generic Viagra	568	$78.80
Cialis	286	$78.00
Cialis/Viagra Combo Pack	172	$74.95
Viagra Super Active+	121	$134.80
Female (pink) Viagra	119	$44.00
Human Growth Hormone	104	$83.95
Soma (Carisoprodol)	99	$94.80
Viagra Professional	87	$139.80
Levitra	83	$100.80
Viagra Super Force	81	$88.80
Cialis Super Active+	72	$172.80
Amoxicillin	47	$35.40
Lipitor	38	$14.40
Ultram	38	$45.60
Tramadol	36	$82.80
Prozac	35	$19.50
Cialis Professional	33	$176.00
Retin A	31	$47.85

(a)

Category	Quantity
Men's Health	1760
Pain Relief	232
Women's Health	183
General Hearth	135
Antibiotics	134
Antidepressants	95
Weight Loss	92
Allergy & Asthma	85
Heart & Blood Pressure	72
Skin Care	54
Stomach	41
Mental Health & Epilepsy	33
Anxiety & Sleep Aids	33
Diabetes	22
Smoking Cessation	22
Vitamins and Herbal Suppliments	18
Eye Care	15
Anti-Viral	14

(b)

Table 3: Table (a) shows the top 18 product items added to visitor shopping carts (representing 66% of all items added). Table (b) shows the top 18 seller-defined product categories (representing 99% of all items).

of time. However, we do not know, on a per-program basis, the actual average purchase price. Thus, we explore three different approximations, all of which we believe are conservative.

First, for on-line pharmacies we use the static value of roughly $100 as reported in our previous "*Spamalytics*" study [10]. However, this study only considered one particular site, covered only 28 customers, and was unable to handle more than a single item placed in a cart (i.e., it could not capture information about customers buying multiple items).

We also consider a second approximation based on the minimum priced item (including shipping) on the site for each program under study. Since sites can have enormous catalogs, we restrict the set of items under consideration as follows. For pharmacy sites, we consider the top 18 most popular items as determined by the analysis of EvaPharmacy in § 4 (these top 18 items constituted 66% of order volume in our analysis). For each of these items present in the target pharmacy, we find the minimum-priced instance (i.e., lowest dosage and quantity) and use the overall minimum as our per-order price. For small deviations between pharmacy formularies (e.g., different Viagra store-brand variants) we simply substitute one item for the other. We repeat this same process for software, but since we do not have a reference set of most popular items for this market, we simply use the declared "bestsellers" at each site (16 at Royal Software, 36 and SoftSales and 76 at EuroSoft)—again using the minimum priced item to represent the average price per order.

Finally, we calculate a "basket-weighted average" price using measured popularity data. For pharmacies we again consider the 18 most popular EvaPharmacy items and extract the overlap set with other pharmacies. Using the relative frequency of elements in this intersection, we calculate a popularity vector that we then use to weight the minimum item price; we use the sum of these weights as the average price per order. Intuitively, this approach tries to accommodate the fact that product's have non-uniform popularity, while still using the conservative assumption that users order the minimum dosage and quantity for each item. Note that we implicitly assume that the distribution of drug popularity holds roughly the same between online pharmacies.[10]

We repeated this analysis, as before, with site-declared best-selling software packages. To gauge relative popularity, we searched a large BitTorrent metasearch engine (isohunt.com), which indexes 541 sites tracking over 6.5 million torrents. We assigned a popularity to each software item in proportion to the sum of the seeders and leechers on all torrents matching a given product name. We then weighted the total prices (inclusive of any handling charge) by this popularity metric to arrive at an estimate of the average order price.

[10]One data point supporting this view is Rx–Promotion's rank-ordered list of best selling drugs. The ten most popular items sold by both pharmacies are virtually the same and ranked in the same order.

Affiliate Program	orders/month	*Spamalytics*		Min product price		Basket-weighted average	
		single order	rev/month	single order	rev/month	single order	rev/month
33drugs	9,862	$100	$980,000	$45.00	$440,000	$57.25	$560,000
4RX	8,001	$100	$800,000	$34.50	$280,000	$95.00	$760,000
EuroSoft	22,776	N/A	N/A	$26.50	$600,000	$84.50	$1,900,000
EvaPharmacy	26,962	$100	$2,700,000	$50.50	$1,300,000	$90.00	$2,400,000
GlavMed	17,933	$100	$1,800,000	$54.00	$970,000	$57.00	$1,000,000
Online Pharmacy	5,856	$100	$590,000	$37.00	$220,000	$58.00	$340,000
Pharmacy Express	7,933	$100	$790,000	$51.00	$410,000	$58.75	$460,000
Royal Software	13,483	N/A	N/A	$55.25	$750,000	$133.75	$1,800,000
Rx–Promotion	6,924	$100	$690,000	$45.00	$310,000	$57.25	$400,000
SoftSales	1,491	N/A	N/A	$20.00	$30,000	$134.50	$200,000

Table 4: Estimated monthly order volume, average purchase price, and monthly revenue (in dollars) per affiliate program using three different per-order price approximations.

5.2 Revenue

Finally, to place a rough estimate on revenue, we multiply the 2011 order volume measurements shown in Figure 4 against each of the previously mentioned approximations, summarized in Table 4. In general, the approximation from our prior "*Spamalytics*" study is the largest, followed by basket-weighted average and then minimum product price. However, for pharmaceutical programs the difference between product prices is not large, and thus the minimum and basket-weighted estimates all lie within 2X of one another. Software programs see much more variation in price, and hence the difference between the minimum and basket-weighted revenue estimates can be substantial.

Using the basket-weighted approximation, we find that both GlavMed and EvaPharmacy produce revenues in excess of $1M per month, with all but two over $400K. Surprisingly, software sales also produce high revenue—less due to high prices than high order volumes. It remains for future work how to further validate how closely order volumes track successfully completed *orders* for this market niche.

5.3 External consistency

While we put considerable care into producing these estimates, a number of biases remain unavoidable. First, while our order volume data has internal consistency (and consistency with order number implementations in common shopping cart software), we could not capture the impact of order declines. Thus, we have a somewhat optimistic revenue estimate, since surely some fraction of orders will not complete.

On the other hand, our estimates of average order revenue are themselves conservative in several key ways. First, they assume that all purchasers select only a single item. Second, they assume that when purchasing an item, all users select the minimum dosage and quantity. Finally, for pharmaceuticals we need to keep in mind that EvaPharmacy does not carry "harder" drugs found at other sites, such as Schedule II opiates. We have found anecdotal evidence that these drugs are highly popular at such sites, but our methodology does not offer any means to consider their impact. Such items are also typically more expensive than other drugs (e.g., the cheapest Hydrocodone order possible at one popular pharmacy is $186 plus shipping). Thus, this other factor will cause us to *underestimate* the true revenue per order.

Our intuition is that such factors are modest, and our estimates capture—within perhaps a small constant factor—the true level of financial activity within each enterprise. However, absent ground truth data for program revenues, it is not generally possible to validate our model and hence verify that our measurements actually capture reality. In general, this kind of validation is rarely possible since the actors involved are not public companies and do not make revenue statements available.

Due to an unusual situation, however, we were able to acquire such information for one program, Rx–Promotion. In particular, a third party made public a variety of information, including multiple months of accounting data, for Rx–Promotion's payment processor.[11] While we cannot validate the provenance of this data, its volume and specificity make complete fabrication unlikely. In addition, given that our research covers only a small subset of this data, it seems further unlikely that any fabrication would closely match our own independent measurements.

Unfortunately, we do not have payment ledgers precisely covering our 2011 measurement period. Instead, we compare against a similar period six months earlier for which we do have ground truth documentation, 27 consecutive days from the end of Spring, 2010. These

[11] While our legal advisers believe that the prior public disclosure of this data allows its use in a research context, we chose not to unnecessarily antagonize the payment services provider by naming them here.

two periods are comparable because during both times Rx–Promotion had significant difficulty processing orders on "controlled" drugs (indeed, during the 2011 period such drugs had been removed from the standard formulary on Rx–Promotion affiliates).[12]

Based on this data, we find that between May 31 and June 26, 2010, Rx–Promotion's turnover via electronic payments was $609K.[13] Using our estimate of 385 orders per day in 2010 (see § 3), this is consistent with an average revenue per order of $58, very similar to our basket-weighted average order price estimate of $57. While we suspect that both estimates are likely off (with the number of true June 2010 orders likely less due to declines, and January 2011 price-per-order likely higher due to conservatism in our approximation), they are sufficiently close to one another to support our claim that this approach can provide a rough, but well-founded estimate (i.e., within a small constant factor) of program revenue.

6 Conclusion

When asked why he robbed banks, Willie Sutton famously responded, "Because that's where the money is." The same premise is frequently used to explain the plethora of unwanted spam that fills our inboxes, pollutes our search results and infests our social networks—spammers spam because they can make money at it. However, a key question has long been how much money, and from whom? In this paper we provide what we believe represents the most comprehensive attempt to answer these questions to date. We have developed new inference techniques: one to estimate the rate of new orders received by the very enterprises whose revenue drives spam, and the other to characterize the products and customers who provide that same revenue. We provide quantitative evidence showing that spam is ultimately supported by Western purchases, with a particularly central role played by U.S. customers. We also provide the first sense of market size, with well over 100,000 monthly orders placed in our dataset alone. Finally, we provide rough but well-founded estimates of per-program revenue. Our results suggest that while the spam-advertised pharmacy market is substantial, with annual revenue in the many tens of millions of dollars, it has nowhere near the size claimed by some, and indeed falls vastly short of the annual expenditures on technical anti-spam solutions.

[12] During periods when such drugs were sold *en masse*, the overall Rx–Promotion revenue was frequently doubled.

[13] Interestingly, this data also provides useful information about refunds and chargebacks (together about 10% of revenue) as well as processing fees (roughly 8.5%). Thus, the gross revenue delivered to Rx–Promotion in June 2010 was likely closer to $489K. Finally, since roughly 40% of successful order income is paid to affiliates on a commission basis, that leaves only $270K (44% of gross) for fulfillment, administrative costs, and profit.

Acknowledgments

We offer our thanks to the many individuals and organizations who aided us in this study. First, we thank both our card issuer and the anonymous provider of the Eva hosting log; together they provided us with the key tools to execute this study. Second, we thank our numerous spam data providers — Jose Nazario, Chris Morrow, Barracuda Networks, Abusix and again as many who prefer to remain anonymous — provided the raw spam data advertising the programs covered in this study. We thank Brian Kantor, Joe Stewart, Kevin Fall, Jeff Williams, Eliot Gillum, Hersh Dangayach and Jef Pozkanzer, among a long list of others, for their operational support and guidance. Erin Kenneally, Aaron Burstein, Daniel Park, Tony Perez and Patrick Schelsinger provided key legal oversight while Kathy Krane, Ellen Sanders, Faye McCullough, Robin Posner, Marianne Generales and Art Ellis provided administrative oversight. We thank Kate Franz for her feedback regarding pharmaceuticals. Finally, we wish to acknowledge the efforts of the anonymous reviewers as well as the feedback and support of the entire CCIED team. This work was supported in part by National Science Foundation grants NSF-0433668, NSF-0433702, NSF-0831138 and CNS-0905631, by the Office of Naval Research MURI grant N000140911081, and by generous research, operational and/or in-kind support from Google, Microsoft, Yahoo, Cisco, HP and the UCSD Center for Networked Systems (CNS). McCoy was supported by a CCC-CRA-NSF Computing Innovation Fellowship.

References

[1] C. Akass. Storm worm 'making millions a day'. `http://www.computeractive.co.uk/pcw/news/1923144/storm-worm-millions-day`, 2008.

[2] M. de Vivo, E. Carrasco, G. Isern, and G. de Vivo. A Review of Port Scanning Techniques. *Computer Communication Review*, 1999.

[3] Forrester Data. Consumer Attitudes Toward Spam in Six Countries. `http://www.bsacybersafety.com/files/Forrester_Consumer_Spam.pdf`, 2004.

[4] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: The Underground on 140 Characters or Less. In *Proc. of 17th ACM CCS*, 2010.

[5] C. Herley and D. Florêncio. A Profitless Endeavor: Phishing as Tragedy of the Commons. In *Proc. of the 11th NSPW*, 2008.

[6] C. Herley and D. Florêncio. Economics and the Underground Economy. Black Hat Briefings, July 2009.

[7] C. Herley and D. Florêncio. Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy. In *Economics of Information Security and Privacy*, 2010.

[8] Ipsos Public Affairs. Key Findings of the 2010 MAAWG Email Security Awareness and Usage Sur-

vey. http://www.maawg.org/system/files/2010_MAAWG-Consumer_Survey_Key_Findings.pdf, 2010.

[9] C. Kanich, N. Chachra, D. McCoy, C. Grier, D. Wang, M. Motoyama, K. Levchenko, S. Savage, and G. M. Voelker. No Plan Survives Contact: Experience with Cybercrime Measurement. In *Proc. of 4th USENIX CSET*, 2011.

[10] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamalytics: An Empirical Analysis of Spam Marketing Conversion . In *Proc. of 15th ACM CCS*, 2008.

[11] S. Karam. Cybercrime is more effective than drug trading. http://www.crime-research.org/news/29.11.2005/1666/, 2005.

[12] Kommersant. Spamming may become criminal offense. http://en.rian.ru/papers/20101202/161593138.html, 2010.

[13] B. Krebs. Spam Affiliate Program Spamit.com to Close. http://krebsonsecurity.com/2010/09/spam-affialite-program-spamit-com-to-close/, 2010.

[14] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage. Spamcraft: An Inside Look at Spam Campaign Orchestration. In *Proc. of 2nd USENIX LEET*, 2009.

[15] LegitScript. Industry Trends: EvaPharmacy, 33Drugs (DrugRevenue) emerge as major Internet threats. http://legitscriptblog.com/2009/10/industry-trends-evapharmacy-33drugs-drugrevenue-emerge-as-major-internet-threats/, 2009.

[16] K. Levchenko, A. Pitsillidis, N. Chachra, B. Enright, M. Felegyhazi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, N. Weaver, V. Paxson, G. M. Voelker, and S. Savage. Click Trajectories: End-to-End Analysis of the Spam Value Chain. In *Proc. of IEEE Symposium on Security and Privacy*, 2011.

[17] Magento. Magento eCommerce Platform. http://www.magentocommerce.com.

[18] S. Malinin. Spammers earn millions and cause damages of billions. http://english.pravda.ru/russia/economics/15-09-2005/8908-spam-0/, 2005.

[19] Marshal. Sex, Drugs and Software Lead Spam Purchase Growth. http://www.m86security.com/newsitem.asp?article=748, 2008.

[20] T. Moore and R. Clayton. An Empirical Analysis of the Current State of Phishing Attack and Defence. In *Proc. of 6th WEIS*, 2007.

[21] Y. Niu, Y.-M. Wang, H. Chen, M. Ma, and F. Hsu. A Quantitative Study of Forum Spamming Using Context-based Analysis. In *Proc. of 14th NDSS*, 2007.

[22] D. Samosseiko. The Partnerka — What is it, and why should you care? In *Proc. of Virus Bulletin Conference*, 2009.

[23] Senate Committee on Commerce, Science, and Transportation. Cybersecurity—Assessing Our Vulnerabilities and Developing An Effective Defense, 2009.

[24] Ubercart. http://www.ubercart.org.

[25] Visa Inc. Visa Check Card Issuer Authorization Performance Self-Diagnostic Tool. http://www.weknowpayments.com/documents/pdf/Visa_Performance_Tool.pdf, 2008.

[26] Y.-M. Wang, M. Ma, Y. Niu, and H. Chen. Spam Double-Funnel: Connecting Web Spammers with Advertisers. In *Proc. of 16th ACM WWW*, 2007.

[27] X-Cart. http://www.x-cart.com.

[28] Zen Ventures, LLC. Zen Cart. http://www.zen-cart.com.