PharmaLeaks: Understanding the Business of Online Pharmaceutical Affiliate Programs

Damon McCoy[◇] Andreas Pitsillidis[*] Grant Jordan[*] Nicholas Weaver[*†] Christian Kreibich[*†]
Brian Krebs[‡] Geoffrey M. Voelker[*] Stefan Savage[*] Kirill Levchenko[*]

[◇]*Department of Computer Science, George Mason University*
[*]*Department of Computer Science and Engineering, University of California, San Diego*
[†]*International Computer Science Institute, Berkeley, CA*
[‡]*KrebsOnSecurity.com*

Abstract

Online sales of counterfeit or unauthorized products drive a robust underground advertising industry that includes email spam, "black hat" search engine optimization, forum abuse and so on. Virtually everyone has encountered enticements to purchase drugs, prescription-free, from an online "Canadian Pharmacy." However, even though such sites are clearly economically motivated, the shape of the underlying business enterprise is not well understood precisely because it is "underground." In this paper we exploit a rare opportunity to view three such organizations—the GlavMed, SpamIt and RX-Promotion pharmaceutical affiliate programs—from the inside. Using "ground truth" data sets including four years of raw transaction logs covering over $185 million in sales, we provide an in-depth empirical analysis of worldwide consumer demand, the key role of independent third-party advertisers, and a detailed cost accounting of the overall business model.

1 Introduction

Much like the legitimate Internet economy, advertising is a major driver for the "underground" criminal economy as well. For all their variety, spam, search-engine abuse, forum spam and social spam—as well as the botnets, fast-flux networks and other technical infrastructure that enable these activities—are all simply low-cost advertising platforms that monetize latent consumer demand. Consequently, an emerging research agenda has developed around understanding the economic structure of these businesses, both to understand the scope and drivers for the problem [8, 9, 13], as well as to help prioritize interventions [14, 15]. Unfortunately, while clever inference and estimation techniques can illuminate a few of the key questions, much remains unclear. This is because, as a rule, there is little "ground truth" data in the field for either validating such results or to provide finer-grained analytics that can be obtained via inference.

This paper provides a rare counter-point to this rule. Under a variety of serendipitous circumstances (largely driven by competition between criminal organizations), a broad corpus of ground truth data has become available. In particular, in this paper we analyze the content and implications of low-level databases and transactional metadata describing years of activity at the GlavMed, SpamIt and RX-Promotion pharmaceutical affiliate programs. By examining hundreds of thousands of orders, comprising a settled revenue totaling over US$185M, we are able to provide comprehensive documentation on three key aspects of underground advertising activity:

Customers. We provide detailed analysis on the consumer demand for Internet-advertised counterfeit pharmaceuticals, covering customer demographics, product selection (including an examination of drug abuse as a driver), reorder rates and market saturation.

Advertisers. We quantitatively detail the role of third-party affiliate advertisers (both email/forum spammers and SEO-based advertisers), the dynamics of their labor market, their ability to drive revenue and the distribution of their commission income. This analysis includes the operators of many of the best-known botnets including MegaD, Grum, Rustock and Storm, and we document individual advertisers generating over $10M in sales.

Sponsors. We derive an empirical revenue and cost model, including both direct costs (sales commissions, supply, payment processing) and indirect costs (hosting, domain registration, program advertisements). We also provide insight and validation about the most significant overheads for the operators of such programs.

This is an unusual research paper. We introduce no new artifact, we develop no new inference technique, we deploy no new measurement infrastructure. We do none of these things because we don't need to; we have the actual data sets that we would otherwise try to measure, infer or estimate. Thus, while there are significant methodological challenges that we must overcome (mainly around the forensic reverse engineering of database schemas and their semantics), ultimately the contribution of this paper is in its results. However, we believe these are both unique and significant, with implications for best addressing this variety of Internet abuse.

2 Background

Abusive Internet advertising has existed virtually as long as the Internet itself. In addition to well-defined advertising channels such as sponsored search [11, 12], rogue advertisers make use of a broad range of vectors to attract customer traffic including email spam [1, 6, 14, 17], search engine manipulation [7, 13, 23], forums and blog spam [19, 24] as well as online social networks [4, 22]. Due to pressure against these tactics, few legitimate merchants will engage such advertisers and thus rogue advertising and rogue products tend to go hand in hand. For example, in one recent report on email spam, Symantec estimated that 80% of all such messages shilled for "prescription-free" pharmaceuticals [21].

However, the structure of this activity has changed significantly over the last decade. In particular, market specialization has largely eliminated the independent "soup-to-nuts" advertiser who previously handled the entirety of the sale process [16]. Instead the rise of the affiliate program, or "partnerka", model has separated the role of the advertiser, paid on commission to attract customer traffic, from the sponsor who in turn handles Web site design, payment processing, customer service and fulfillment [18]. This evolution is not unique to abusive advertising; indeed, large legitimate merchants such as Amazon also sponsor affiliate programs as a means of advertising. However, it has been deeply internalized within the underground ecosystem including the pay-per-install [3], FakeAV [20], pornography [25], pharmaceuticals [2], herbal supplements [14], replica [14] and counterfeit software markets [9], among others.

Counterfeit pharmaceuticals represent a typical example. Here a range of sponsoring affiliate programs provide drugstore storefronts, drug fulfillment (typically via drop shipping from India), payment processing, customer service and so on. Independent advertisers, or *affiliates*, in turn promote the program (e.g., by using botnets to send spam email or manipulating search engine results) and are paid a commission on each sale that results from a click on one of their ads. Commissions range from 30%–40% of gross revenue, typically paid via a quasi-anonymous online money transfer service such as WebMoney or Liberty Reserve.

This business model has two key advantages for the advertiser: focus and mobility. Without needing to attend to issues such as Web site design, payment processing, customer service, fulfillment and so on, the advertiser is free to focus single-mindedly on the task of attracting customer traffic to these sites. Indeed, this functional specialization has supported the creation of ever more sophisticated botnets for email delivery or "black hat" search engine optimization, and many of the largest botnets are directly involved in advertising the programs in this paper (Rustock, MegaD, Grum, Cutwail, Storm, Waledac and others). The second advantage of this model, mobility, is that the loosely coupled nature of their relationship with affiliate programs allows an advertiser to switch programs at will (or even support multiple programs at once). This low "switching cost" provides bargaining power for the effective advertiser (indeed, we witness high-sales advertisers able to use this threat to drive higher commissions). More importantly, it reduces an advertiser's exposure to business continuity risk. If a particular affiliate program should shut down, advertisers can still monetize their investments (e.g., in a botnet) by advertising for a different sponsor.

However, the benefits of this separation are strong for the sponsoring affiliate program as well. By outsourcing advertising they free themselves from direct exposure to the criminal risks associated with large-scale advertising enterprises (e.g., mass compromise of computers and online accounts). Second, because advertisers are paid on a commission basis, they also outsource "innovation risk". Program sponsors need not predict the best way to attract customer traffic at a given point in time. Instead hundreds of advertisers innovate independently; if many of them fail, so be it. Since advertisers are only paid commissions on successful sales, a sponsor will only end up paying for effective advertising strategies and need not distinguish among strategies *a priori*.

Against this background, online pharmaceutical sales is one of the oldest and largest affiliate program markets. This market supports tens of affiliate programs and, as we will see, thousands of independent advertisers (*affiliates*) and hundreds of thousands of customers. However, while the mechanics of this business model are well-described in recent work [2, 14, 18], the dynamics of the actors and the underlying constants that define the cost structure (and hence the vulnerabilities in the business) are not well understood at all. Indeed, even simple questions such as "How big is sales turnover?" are imperfectly understood. For example, Kanich *et al.* used one method to estimate that the combined turnover across seven leading pharmacy programs (constituting two-thirds of affiliate brands advertised in spam) is roughly 86,000 orders per month [9]. However, Leontiadis *et al.* use a different technique to arrive at a much larger estimate suggesting over 640,000 orders per month [13].

In this paper, we answer this and many other such questions *precisely* by focusing in depth on three pharmaceutical affiliate programs: GlavMed, SpamIt and RX-Promotion. These organizations have been in business for five years or more. Together, they represent many tens of storefront "brands" (including the ubiquitous "Canadian Pharmacy") and, according to the data from our prior measurement studies, these programs have been advertised in over a third of all spam email messages [14].

3 Authenticity and Ethics

Our use of "found data" creates two new concerns that we address here: authenticity and ethics.

First, it is useful to provide some rough context concerning the circumstances leading to the release of these data sets. As explained in the previous section, GlavMed and RX-Promotion are both long-operating pharmaceutical affiliate programs based in Russia. However, for a variety of reasons, enmity developed between owners in each program, revealed anecdotally through "sniping" on underground forums, claims of denial-of-service attacks and ultimately to the hacking of each other's infrastructure sites. Perhaps inspired by the "online leak" meme, popularized recently by Wikileaks and others, elements of these two organizations (or parties sympathetic to their positions) gained access to information about each other's operations and then made portions of this data available: sometimes publishing very broadly on underground forums and file-sharing sites, and other times distributing to a variety of journalists, e-crime researchers, law enforcement agencies as well as a broad range of underground actors.

Through these channels we obtained access to three transactional data sets: the complete dump, covering four years, of the GlavMed and SpamIt back-end database (comprising transactions, payments and so on) and a year of more restricted transactional data for the RX-Promotion program. We also received two metadata corpuses: detailed archived chat logs from the program operator for sites operated by GlavMed and SpamIt, as well as financial data concerning the revenue and cost structure for the RX-Promotion program. For further context and back-story about this data, we refer readers to the "Pharma Wars" series by Brian Krebs [10].

3.1 Authenticity

Given that we did not gather the information ourselves *and* the adversarial nature by which the data became available, an obvious question is how to evaluate its accuracy and authenticity: how do we know that our sources did not fake the data?

While we cannot establish clear provenance beyond all possible doubt, we observe a range of strong supporting evidence. First, we observe that the data sets are large and detailed (over 2M sales records, with over 140 linked tables, coupled with several GB of related metadata). These attributes do not entirely discount the possibility that they could be grossly fraudulent, but it suggests that the costs of creating such a forgery would be significant.

Second, we consider questions of internal and cross-consistency. The transactional data sets have complex schemas (covering orders, potentially many payment transactions per order, commissions to advertisers, subsequent payouts, and so on) and we find direct concordances between the different elements (e.g., if we sum the settled sales for a particular affiliate it typically relates directly to the size of the payout to that affiliate). We also find concordances *between* the transactional data and the metadata. For example, we found multiple chat logs directing a GlavMed/SpamIt employee to make a payment to a particular affiliate that is then matched by an identical payout record in the associated transactional database. Similarly, the monthly revenue for shipped products for RX-Promotion is consistent with the settled revenue from its payment processor in the same period. Finally, during the period covered by all three transactional data sets we had placed multiple product orders from each of the associated programs [9, 14]. We find each and every one of our orders in the appropriate database with the correct data.

While this evidence cannot comprehensively prove the absence of fraud,[1] given the strong concordances and the absence of any evidence supporting the forgery hypothesis, we believe the greater likelihood is that these data sets are authentic and accurate. We proceed with this assumption going forward.

3.2 Ethics

The other fundamental issue concerns the ethics of using data that was, in all likelihood, gathered via illegal means. Here there are two kinds of questions. The first is a high-level question concerning whether the nature of how the data was originally gathered should *prima facie* proscribe all subsequent uses of it. This question is not new and it manifests in a range of fields. For example, should a political scientist be proscribed from analyzing the contents of the Pentagon papers (or the more contemporary Wikileaks data) in reasoning about U.S. foreign policy? Similarly, should researchers avoid using widely publicized stolen password data (e.g., from the Anonymous/Lulzsec leaks) when studying the strength of user-selected passwords? We justify our own choice to take such steps by reasoning about harm.

We observe that this data is already broadly available and the knowledge of its existence, its association with the GlavMed, SpamIt and RX-Promotion organizations, and some of the over-arching contents (e.g., total revenue, etc.) have already been widely and publicly documented. Consequently, we cannot create any new harm simply through association with these entities or repeating these findings.

To manage any remaining harms we institute a number

[1] For example, while we believe comprehensive forgery would have been cost prohibitive given the size and richness of these data sets, a forger might have selectively altered only certain records and updated dependent schemas to be consistent.

Program	Period	Affiliates	Customers	Billed orders	Revenue
GlavMed	Jan 2007 – Apr 2010	1,759	584,199	699,516	$81M
SpamIt	Jun 2007 – Apr 2010	484	535,365	704,169	$92M
RX-Promotion	Oct 2009 – Dec 2010	415	59,769 – 69,446	71,294	$12M

Table 1: Summary of the affiliate program data used in the analysis. Orders are rounded to the nearest thousand, revenue to the nearest million U.S. Dollars. Affiliates and customers are listed *after* de-duplication and billed orders and revenue reflect only those orders whose payment transactions completed (both processes are described in Section 4.1).

of controls in our work focused on the individual stakeholders. First and foremost, and in accordance with our institution's human subjects review process, we protect customer confidentiality since, of all parties described in the data, they are most vulnerable. To this end, we committed to modify the raw data sets to anonymize personally identifiable customer data such as their name, address and the PAN component of their credit card information (though in a way that we are able to associate multiple orders from the same customer). For the remaining stakeholders, program employees, affiliates, suppliers and payment processors, we use a similar standard in publishing our work. In each of these cases the persons or organizations operate using handles or code names that are not clearly identifiable (e.g., "brainstorm" or "gl") without the use of additional data sources. In some cases (e.g., payment processors, suppliers) we have become aware of the likely true names of these organizations (typically through reading the metadata) but we restrict ourselves to using these non-identifiable code names since the true names do not enhance our analysis. We do not name program employees and we typically discuss affiliates in aggregate, with an exception being the top affiliates whom we distinguish in this paper using only their online handles.

4 Derived Data

Using "found data" also introduces a range of methodological challenges, ranging from reverse engineering schemas to resolving ambiguities in the data. In this section we describe the data sets (summarized in Table 1) and explain how we derived the additional contextual relations used in our analysis.

4.1 GlavMed and SpamIt

The first two data sets are PostgresSQL database dumps of the operational databases for the GlavMed and SpamIt programs, including all schemas, data, and trigger functions, but no other code external to the database. The GlavMed database begins November 2005 and ends early May 2010, of which we use the period spanning all of 2007–2009 and the first four months of 2010.[2] GlavMed and SpamIt are sister programs run by the same organization and, indeed, both use the same database schema. In fact, it appears that SpamIt was "forked" from the GlavMed database on June 19, 2007: all records before that date are identical in both databases, while records after that date are distinct. Leaked chat logs of the program operators suggest that this split was related to the owner's contemporaneous acquisition of `Spamdot.biz`, a popular closed spammer forum of that period. In part through this forum, the SpamIt program nominally catered to a select group of affiliates relying on email and other forms of spam, while GlavMed remained open to a broader range of advertisers who primarily advertised via search engine optimization techniques.[3]

A detailed description of the data and its associated schema, consisting of over 140 tables in each database, is outside the scope of this paper. However, we perform most of our analysis using five tables: `shop_sales` describing each order, `shop_transactions` recording attempts to bill (or refund) the order via a payment service provider, `shop_customers` recording customer information, `shop_affiliates` recording information about each affiliate, and `shop_affiliates_income_2` recording affiliate commissions for each sale. We also relied on instant message chat logs of the operators of GlavMed and SpamIt to aid our understanding and validate our hypotheses about the meaning and use of various tables.

However, the GlavMed and SpamIt databases are fundamentally operational in nature, and not naturally designed for the kind of broad analysis that are the goal of this paper. Thus, we now describe the additional data processing required to produce necessary relations (e.g., such as identifying unique customers).

4.1.1 Customers

In an ideal world, each customer record would represent a unique customer and include accurate demographic information for our analysis (age, sex, and either country or U.S. ZIP code). The reality, hardly unique to our data set, is less obliging: In addition to many test accounts

[2] Since our goal is accuracy and not completeness, we purposely exclude the first 14 months of the data set because it is both "poisoned" with transactions for other kinds of products, including $500K in counterfeit software sales, and makes inconsistent use of the database schemas that become standard in the later portion of the date range.

[3] This distinction is not absolute, however; domains advertised by GlavMed affiliates have appeared in email spam.

used by the store operators, a large number of customer records are generated by irate users venting their frustration with the deluge of spam advertising the program.[4] Thus, for the purpose of this study, we consider only customers who have successfully placed an order (more specifically, those whose credit card or other payment mechanism was successfully billed, as described later), which reduces the number of customer records by 21% in the GlavMed data set (from 875,457 to 690,590) and 39% in the SpamIt data set (from 1,145,521 to 693,319), the latter clearly attracting more abuse.

De-duplication. An additional problem is that, unless the customer uses a previously assigned customer number to explicitly log in, each repeat order would result in a new customer record. To identify repeat customers, we de-duplicate the remaining customer records by coalescing those whose name, billing address and email address are identical, reducing the number of unique customers to 584,199 in GlavMed and 535,365 in SpamIt. For address matching, we used the common Visa/MasterCard Address Verification System (AVS) predicate, which relies on street number and ZIP code only. Both names and email address matches were case insensitive, and we allowed first and last names to be transposed.

Demographics. Our analysis relies on customer demographic data consisting of the customer's country or U.S. ZIP code, as well as their self-reported age and sex. The country and ZIP code are necessary for proper order fulfillment, and therefore are generally reliable. However, customers optionally provide age and sex data when ordering, so it is not always present and it is subject to misreporting. Only 41% of GlavMed orders and 38% of SpamIt orders included this information, and we cannot validate it since customers could easily dissemble. Indeed, we found that a larger than expected number of users reported birth dates of January 1, February 2, and so on (these being some of the easiest dates to report via the interface). However, these anomalies are a small minority and we proceed under the assumption that the data is generally correct (eliminating these cases does not substantively change the results reported in Section 5.1.3).

4.1.2 Affiliates

As with customers, affiliate records also require de-duplication. However, here the duplication is not a mere artifact of the interface, but is frequently an intentional action. Affiliates frequently register under multiple identities, either to modulate their perceived earnings (affiliate programs commonly provide "top" lists showing the affiliates with the highest earned commissions) or to gain access to additional referral commissions that are provided on sales generated by new affiliates referred into the program.[5] To address these issues, we de-duplicate affiliates as follows. For all affiliates with over $200 in revenue we link those who share an email address, ICQ number[6] or "identified commission payments". We considered a commission payment to be identified if it represents over 75% of an affiliate's revenue and includes unique payment account information (such as a WebMoney, Fethard Finance, or ePassporte account or an identified GlavMed payment card). The notion of identified payments was necessary to avoid incorrectly associating affiliates who use the commission payments system to pay third parties (e.g., by asking for small payouts to a third-party WebMoney purse).

4.1.3 Transaction Outcomes

In the GlavMed and SpamIt data sets, each customer sales record in turn drives the creation of one or more transaction records which reflect an attempt to transfer money to or from a customer (as identified by a credit card or Automated Clearing House (ACH) identifier) via a third-party payment service provider. When a transaction is successful the `response_status` field in this record is zero (we validated these semantics by examining both raw payment processing error messages and associated SQL triggers in the databases).

However, for a host of reasons transactions are frequently declined. Indeed, over 25% of all transaction attempts decline in both the GlavMed and SpamIt data sets. In these cases, new transactions may be generated, possibly using different payment service providers. In some cases, large order amounts are billed into two smaller transactions. Overall, 91% of sales are able to complete a payment transaction.

Finally, a transaction may be refunded, either partially or fully. An additional complexity arises from currency conversion because customer payments are internally valued in U.S. Dollars, but can arrive in Euros, Pounds and several other currencies. When refunds arrive in native currency, we locate the original transaction and calculate the dollar refund value on a pro-rated basis against the original value in the native currency. All revenue numbers reported in the analysis refer to the total amount billed, before any refunds against the transaction. Refunds are shown separately in Table 3.

Note that having this ground truth data allows us to calibrate biases in previous methods for estimating revenue. In particular, we revisit our "purchase pair" tech-

[4] This frustration was well captured by the many regular expressions in the operators' customer blacklist, e.g., `(.*)SP(A+)M(.*)` and `(.*)F(U+)CK(.*)`.

[5] As an incentive to attract affiliates, program sponsors will typically offer their affiliates a 5% commission on the future sales of any new affiliate they bring into the program.

[6] ICQ is one of the oldest widely-deployed IM chat systems, and is very popular in Russia and CIS states.

nique that infers order turnover via customer order number advancement and then conservatively estimates the average order size to gauge overall revenue [9]. Across four years, we find that a significant number of order numbers never appear in the database due to either filtering for customer fraud or shopping cart abandonment (between 13–28% for SpamIt and 7–17% for GlavMed). The lower number of absent orders for GlavMed is likely because the search engine vector used by its affiliates generates less antipathy among consumers. In both cases, 8–12% of the orders that do appear in the database are ultimately declined and do not ship. Consequently, true turnover is between 8% (low of GlavMed) and 35% (high of SpamIt) less than predicted by the "purchase pair" technique. However, since the average successful order size is between $115 (GlavMed) and $135 (SpamIt), revenue estimates basd on an average sale of $100 are roughly in-line with true revenue (within 6% overall for GlavMed and 13% overall for SpamIt).

4.2 RX-Promotion

Our third data set concerning transactions from the RX-Promotion program is far more limited. It only covers a single year of data from January to December of 2010, consisting of a single extracted view summarizing each sale during the period *made by U.S. customers*. In addition, roughly one week of data is missing (around the last week of April 2010). Consequently, this transactional data will strictly understate the turnover from RX-Promotion.[7]

Each sales record includes information about the customer (name only), the status of the order, its contents, the total price as well the amount paid to the supplier, shipper and the affiliate who generated the sale. Our analysis includes only orders with the status value "shipped", which make up 77% of all sales records ("declined" was the next largest category at 14%).

Since the RX-Promotion data set does not include crisp customer identifiers, we use two approximations for identifying multiple orders belonging to the same customer. The conservative approximation of 69,446 customers only links sales records together if a customer explicitly logs into the site using a previously assigned customer ID. However, we note that this measure strictly overestimates the number of customers since many users prefer to place subsequent orders by entering in their information again. Alternatively, one can group customers that share the same first and last name (normalized for capitalization), resulting in 59,769 customers. This approach will accurately capture multiple orders from the same user, but at the expense of potentially aliasing users who happen to share the same first and last names. Thus, the true number of unique customers is likely between the two estimates, but to avoid aliasing issues we use the larger conservative estimate in our analyses.

[7] Based on our measurements of both the GlavMed and SpamIt data sets, our own previous study of the Eva Pharmacy program [9], and inference from the RX-Promotion metadata, we are confident that U.S. customers represent between 75% and 85% of total turnover. In addition, the missing week of data from April should cause our data to underestimate annual orders by an additional 2%.

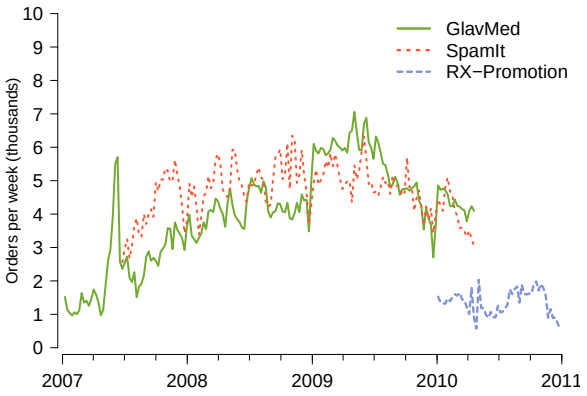


Figure 1: Weekly sales volume for each of the programs.

Finally, we also make use of seven months of overlapping metadata that includes detailed spreadsheets accounting for month-by-month costs and cash flow. This data does not have any of the previous limitations and captures the financial performance of the program precisely and in its entirety.

5 Analysis

Using these data sets, we now provide a detailed assessment of the affiliate program business model. From the standpoint of the program sponsor, we consider four key aspects of the business enterprise in turn: customers, affiliate advertisers, costs and payment processing.

5.1 Customers

Neither online pharmacies nor their advertisers generate capital on their own. These activities thrive only because they exploit latent customer demand for the products on offer. It is this customer purchasing that drives the entire ecosystem and thus this is where we begin: how many purchases, for what, by whom and, perhaps, why?

Overall, as shown in Table 1, 584,199 unique customers placed orders via GlavMed during the measurement period and 535,365 placed orders via SpamIt; of these approximately 130K appear in both. RX-Promotion is a smaller program and covers a shorter time period, with somewhere between 59,769 and 69,446 distinct customers placing orders. In turn these customers generated almost 1.5M orders, varying from week to week as shown in Figure 1. Note that the spike in May 2007 for GlavMed is an artifact corresponding to the short period after GlavMed had purchased SpamIt, but before they

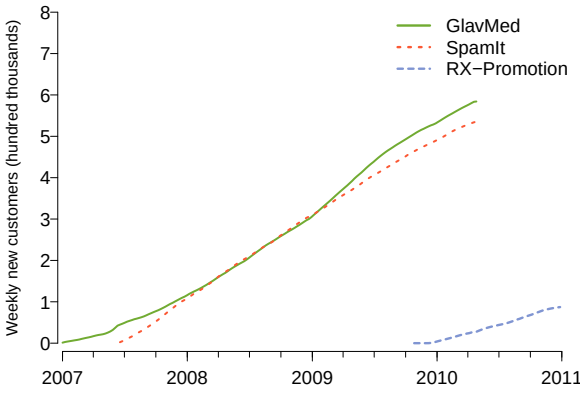


Figure 2: Cumulative number of new customers.

had forked the databases in June 2007 (Section 4.1). After the fork, GlavMed has very steady growth in orders until mid-2009, even surpassing SpamIt, and then starts to decline. Orders to SpamIt plateau for 2008–2009, similarly declining in mid-2009.[8] RX-Promotion order volumes are considerably more dynamic, for reasons we will explain later, with totals varying between 1–2 thousand per week across the year of data.

5.1.1 First-time Customers

However, these million plus customers and their purchases do not necessarily constitute the entirety of this market, but only the *portion* that has been serviced to date by these particular programs. This raises the question: How saturated is the market for counterfeit pharmaceuticals? To evaluate this, Figure 2 shows the cumulative number of unique customers seen in each program per week over the measurement period. Thus, changes in slope indicate changes in the rate of new customer acquisition. From these trends it is clear that that the affiliate programs are attracting new customers at a steady rate over time, and that the market *does not appear to be saturating at all*. In particular, sister programs GlavMed and SpamIt attract new customers at nearly the same rate (3,367/week and 3,569/week on average) while RX-Promotion, a smaller program, attracts customers at a slower, but still constant rate (1,429/week on average). The stability of this growth over time provides some explanation for why spammers continue to blast email indiscriminately to all Internet users over time: they are still mining a rich vein of latent customer demand.

[8]This decline undoubtedly has many roots including increasing pressure that mounted on SpamIt due to its high visibility (e.g., the principal owner of SpamIt was identified by Russian Newsweek as the World's Biggest Spammer), shutdowns of large botnets operating as affiliates (e.g., the MegaD botnet, which we observed spamming for sites associated with SpamIt affiliate "docent", ceased operating in November of 2009), and inter-program competition (e.g., starting in 2010, we see a roughly 15% reduction in the number of active affiliates in the SpamIt program and we witness one large affiliate, "anonymouse", leaving SpamIt and moving to RX-Promotion during this period).

5.1.2 Repeat Customers

New customers, however, are not the whole story. The graphs in Figure 3 show total program revenue per week broken down into two components: revenues from first-time customers and revenue from repeat orders from existing customers. What we see is that repeat orders are an important part of the business, constituting 27% and 38% of average program revenue for GlavMed and SpamIt, respectively. For RX-Promotion revenue from repeat orders is between 9% and 23% of overall revenue.

Overall, revenue from repeat customers steadily increases over the years for GlavMed and SpamIt, and holds steady even when orders and overall revenue decline in mid-2009. The situation is more dynamic for RX-Promotion with a pronounced dip in program revenue in the middle of 2010 that impacts new and repeat customers both. This dip corresponds to the period when RX-Promotion lost its payment processing services for scheduled drugs.[9] Indeed, if we only consider the period after August 2nd, repeat order revenue averages between 12% and 32%.

This data highlights a counterpoint to the conventional wisdom that online pharmacies are pure scams: simply taking credit cards and either never providing goods or providing goods of no quality. Were this hypothesis true, we would not expect to see repeat purchases—clear signs of customer satisfaction—in such numbers. Anecdotally, we have placed several hundred such orders ourselves and, while we cannot speak to the quality of the products we received, we have almost always received a product in return for our payment [9, 14].

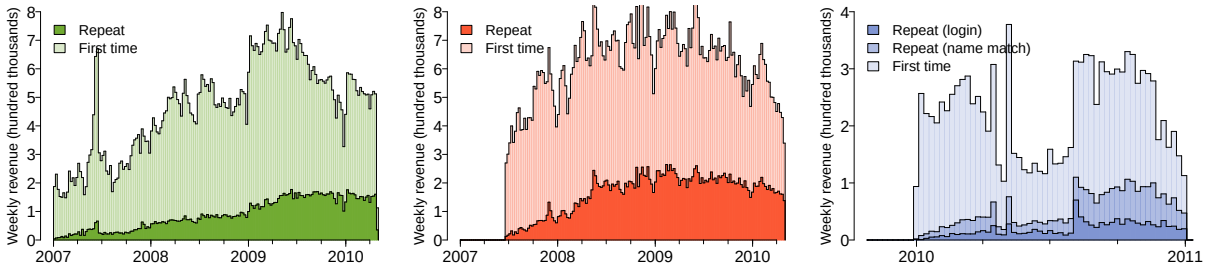
5.1.3 Product Demand

Beyond measuring overall demand, we are particularly interested in determining what makes up this demand: which drugs are being purchased, and does this provide clues about *why* this market is preferred.

In an effort to reach all customer niches, each of the programs carries thousands of products. To reason about this multitude of drugs, we classified the bulk of the products into broad categories based on our best assessment (necessarily subjective) of the drug's use: erectile dysfunction, pain/inflammation, male enhancement (not ED), mental health, sleep, obesity and other.

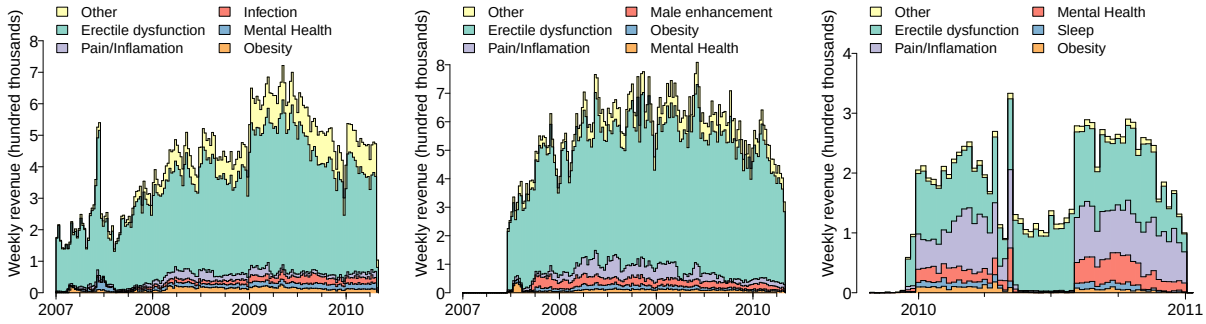
Using this classification, customer demand for specific kinds of drugs in the different programs is striking. As with the previous time series graphs, Figure 4 shows weekly revenue for the three affiliate programs over time,

[9]Associated metadata suggests that RX-Promotion's payment service provider (PSP) had arranged for merchant accounts at an Icelandic bank to be used for RX-Promotion controlled drug payments. However, on May 10th 2010, a complaint by Visa caused the bank to shut down these accounts and thus processing for controlled substances was curtailed until August 2nd when the PSP established new accounts for this purpose with Azeri banks.



(a) GlavMed (b) SpamIt (c) RX-Promotion

Figure 3: Weekly order revenue shown by customer class.



(a) GlavMed (b) SpamIt (c) RX-Promotion

Figure 4: Weekly order revenue shown by drug type.

but here each of the top five revenue-earning drug categories is colored distinctly. For GlavMed and SpamIt, the jokes about spam are spot on: "erectile dysfunction" (ED) purchases dominate their revenue. Customers do purchase other notable drugs, but they represent a small fraction of revenue over time for these programs.

In contrast, revenue from pain/inflammation orders matches revenue from ED in RX-Promotion. RX-Promotion has a markedly different formulary from GlavMed and SpamIt, prominently offering products that GlavMed and SpamIt do not sell. Specifically, these include scheduled drugs for pain (Oxycodone, Hydrocodone, Vicodin, etc.), mental health (Adderal, Ritalin, etc.), and sleep (Valium, etc.), all of which have high abuse potential.[10]

These examples suggest that there may in fact be a range of distinct reasons *why* different drugs are popular via this medium. Table 2 summarizes order volume and program revenue for different groups of drugs sold to customers by the three affiliate programs. Here we merge our original set of categories into three groups that correspond to different customer motivations for purchasing drugs. The first group includes erectile dysfunction (ED), male enhancement, and related products (including fakes such as "Herbal Viagra"). These drugs, sometimes called "lifestyle" drugs, do not address chronic or acute illness. While they are relatively easy to obtain under prescription, seekers may prefer the online channel for reasons of embarrassment or price.[11] The second group includes drugs that have the potential to be seriously abused, and includes addictive drugs such as opiates, depressants, stimulants, etc. For many of these drugs, customers run substantial legal risk in purchasing them without prescription, and presumably run this risk because of a strong desire or need. The third group includes drugs for treating chronic or acute illnesses. Since these drugs carry no strong abuse risk, nor represent a clear cause for social discomfort, we presume that their purchase is motivated by economics: lower direct drug costs (which can be substantial) and the absence of indirect costs (for a doctor's visit). In each category, the table also lists the top categories or specific products.

Reflecting Figure 4, the ED group dominates items ordered and revenue to the program, particularly for GlavMed and SpamIt. For RX-Promotion, though, drugs with the potential for abuse are high-revenue orders. Although they comprise just 14% of orders for

[10]The Controlled Substances Act in the U.S. defines five drug "schedules", or classifications, according to various criteria such as potential for abuse. Scheduled drugs require prescriptions and have heavy financial and/or criminal penalties for illegal sale.

[11]The per-item drug price offered by such programs is frequently less than 20% of that offered by legitimate retailers. For example, the median price for 10 tablets of 100mg Sildenafil Citrate was $42.57 on GlavMed and $23.40 at RX-Promotion. By contrast, according to data at `drugs.com`, legitimate brand Viagra in the same amount sells for $193.99. Note that these prices do not account for shipping, which can add $15 to $30 per order.

	GlavMed		SpamIt		RX-Promotion	
Product	Orders	Revenue	Orders	Revenue	Orders	Revenue
ED and Related	580K (73%)	$55M (75%)	670K (79%)	$70M (82%)	58K (72%)	$5.3M (51%)
Viagra	300K (38%)	$28M (38%)	290K (34%)	$31M (36%)	33K (41%)	$2.7M (27%)
Cialis	180K (23%)	$19M (26%)	190K (22%)	$23M (27%)	18K (22%)	$1.9M (19%)
Combo Packs	49K (6.1%)	$3.9M (5.4%)	110K (14%)	$8.4M (9.8%)	5100 (6.4%)	$350K (3.4%)
Levitra	32K (4.1%)	$3.2M (4.4%)	35K (4.2%)	$3.9M (4.5%)	1200 (1.5%)	$150K (1.5%)
Abuse Potential	48K (6.1%)	$4.5M (6.1%)	64K (7.6%)	$6.2M (7.3%)	11K (14%)	$3.3M (32%)
Painkillers	29K (3.7%)	$2.4M (3.3%)	53K (6.3%)	$4.7M (5.5%)	10K (13%)	$3.0M (29%)
Opiates	—	—	—	—	8000 (10%)	$2.7M (26%)
Soma/Ultram/Tramadol	20K (2.5%)	$1.8M (2.4%)	46K (5.5%)	$4.1M (4.8%)	1000 (1.3%)	$150K (1.5%)
Chronic Conditions	120K (15%)	$9.5M (13%)	64K (7.6%)	$5.2M (6.1%)	8500 (11%)	$1.3M (13%)
Mental Health	23K (2.9%)	$2.1M (2.9%)	16K (1.9%)	$1.4M (1.7%)	6000 (7.4%)	$1.1M (11%)
Antibiotics	25K (3.2%)	$2.1M (2.9%)	16K (1.9%)	$1.4M (1.6%)	1300 (1.6%)	$97K (0.9%)
Heart and Related	12K (1.5%)	$770K (1.1%)	9700 (1.2%)	$630K (0.7%)	390 (0.5%)	$35K (0.3%)
Uncategorized	48K (6.0%)	$4.0M (5.5%)	47K (5.6%)	$3.9M (4.6%)	2400 (3.0%)	$430K (4.2%)

Table 2: Product popularity in each of the three programs. Product groupings and categories are in italics; individual brands are without italics. Opiates are a further subcategory of Painkillers, and include Oxycodone, Hydrocodone, Vicodin, and Percocet. Note, this table *only* describes revenue from drugs and does not capture shipping charges, which are orthogonal to drug popularity.

RX-Promotion, they account for nearly a third of program revenue, with the Schedule-II opiates—only available at RX-Promotion—accounting for a quarter of revenue. Indeed, during the period when RX-Promotion had working credit card processing for controlled meds, sales of Schedule II, III and IV drugs produced 48% of all revenue! The fact that such drugs are over-represented in repeat orders as well (roughly 50% more prevalent in both RX-Promotion and, for drugs like Soma and Tramdol, in SpamIt) reinforces the hypothesis that abuse may be a substantial driver for this component of demand.

5.1.4 Demographics

Although ED drugs account for the majority of business for affiliate programs, focusing on the remaining products reveals remarkably pronounced age and sex trends among customers.

Focusing on customers reporting age and sex information, Figure 5 shows the percentage of all items ordered as a function of age, sex, and detailed product category for GlavMed and SpamIt (excluding ED products, as they would overwhelm the graph). The left half of each graph shows results for women, and the right half shows results for men. The *y*-axis is the self-reported age of customers, and the *x*-axis is the percent of all items these customers ordered. For each age the graphs show stacked horizontal bars, with segments for the top ten *non-ED* product categories.

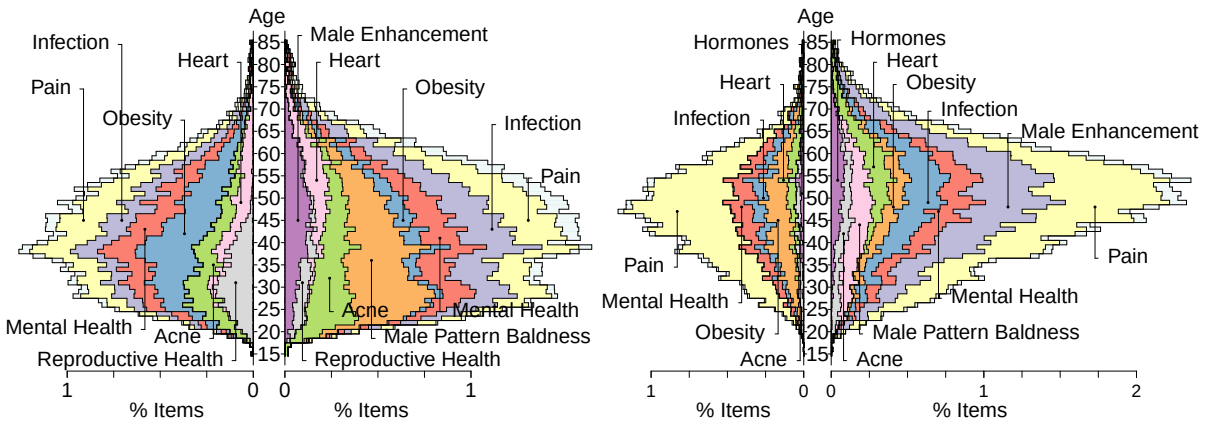
Both age and sex purchasing patterns emerge from this visualization. For example, male GlavMed customers in Figure 5(a) purchase male pattern baldness products (peaking between ages 20–30) and male enhancement products (peak 45–50), while women predominantly purchase obesity (peak 40–45) and reproductive health products (peak 25–30).[12] Mental health and pain/inflammation products are roughly equally popular for men and women, with an older age bias for men.

In contrast to GlavMed, just a few categories predominate for SpamIt in Figure 5(b): pain/inflammation, infection, and mental health for both men and women, male enhancement for men. Other categories more popular in GlavMed, such as acne and male pattern baldness, are smaller. One explanation is that the differences in product popularity correlates with the vector used to advertise the different affiliate programs. Since GlavMed is more likely to be involved in search engine optimization (SEO) oriented advertising, they have an opportunity to target narrower markets (e.g., by manipulating search results for keywords correlated with specific product categories). By contrast, spam is an indiscriminate advertising medium and customers clicking on spam-advertised links are predominantly taken to storefronts advertising ED products. Thus, for these customers to buy other products would require additional initiative to search within the site.

5.1.5 Geography

While both affiliate programs are located in Russia, most of their customers are not. Based on customer shipping addresses, we can determine that, across GlavMed and SpamIt programs, customers from the United States dominate at 75% of orders, with Canada, Australia, and populous countries in Western Europe following in single digits. Emphatically, Western money fuels these af-

[12] Interestingly, male customers also purchase the estrogen drug Clomid, which we have come to understand may be explained by body builders who commonly abuse the drug to counter some of the side-effects of steroids.



(a) GlavMed (b) SpamIt

Figure 5: Items purchased separated into product category and customer age. The left half of each graph show orders from women, and the right half shows orders from men. Customers restricted to those who self-report age and sex.

filiate programs with the U.S., Europe, Canada and Australia constituting 97% of all orders, consistent with the breakdown previously observed in [9].[13]

5.2 Affiliates

While customer purchasing drives the online pharmaceutical ecosystem, affiliates are the ones who attract and deliver the customers—and their money—to the online pharmacies. Affiliates operate by commission, receiving a significant fraction (typically 30–40%) of each customer purchase that reflects the substantial risk they assume in their aggressive advertising activities. Next we analyze the role affiliates play in making online pharmaceutical programs successful as a business.

As discussed in Section 4.1.1, we merge separate accounts in the GlavMed and SpamIt databases that belong to the same affiliate. After account merging, during the 2007–2010 measurement period 1,037 affiliates were active in GlavMed and 305 in SpamIt. Lacking detailed account profile information in RX-Promotion, we consider each account a separate affiliate. With this assumption, during the smaller one-year period for RX-Promotion 415 affiliates were active.

5.2.1 Program Revenue

GlavMed and RX-Promotion are open affiliate programs, and as such they actively advertise and recruit new affiliates to join their programs (with the public advertising focused on SEO-based advertising vectors). SpamIt, on the other hand, is a closed program—focused specifically on email spam—where affiliates join by invitation (Section 4.1). These models influence the kinds of affiliates in a program, the impact they have on generating revenue for a program, as well as the commissions they earn.

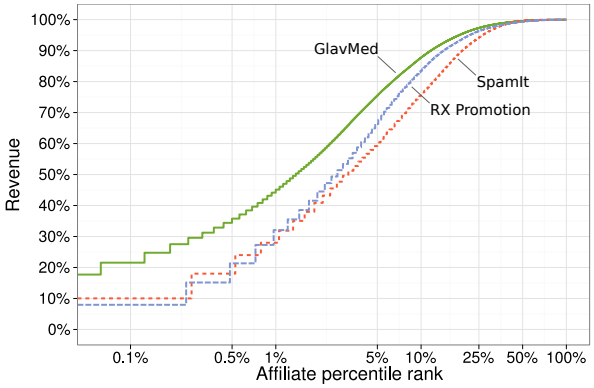


Figure 6: Distribution of affiliate contributions to total program revenue for each program.

Although the programs contain hundreds to thousands of affiliates, most affiliates contribute little to the overall revenue of the programs. Figure 6 shows the CDFs of affiliate contributions to total program revenue for the three affiliate programs. The x-axis is the percent of affiliates, sorted from highest to lowest revenue they generate for the program, and the y-axis is the percent of total program revenue. The graph shows that just 10% of the highest-revenue affiliates account for 75–90% of total program revenue across the three affiliate programs; for GlavMed and RX-Promotion in particular, the remaining 90% of affiliates bring in just 10–15% of total revenue.

In the end, the most important affiliates for a program are just a small fraction of all affiliates. From a business perspective, programs can focus their attention and en-

[13] This previous study also identified substantive differences in the make-up of drugs purchased in the U.S. vs. other Western countries (with U.S. customers driving a disproportionate fraction of demand for non-ED meds). While we still observe this pattern in the SpamIt data (with the fraction of non-ED meds in U.S. customer orders being 3.8× larger than for Europe and Canada), it is absent in GlavMed customers, suggesting that the advertising vector plays a key role in this effect.

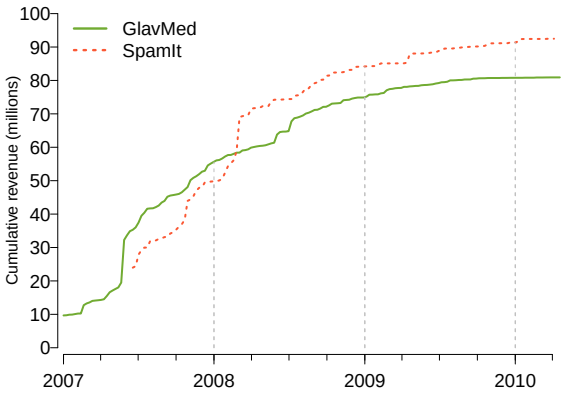


Figure 7: Cumulative contribution of new affiliates over time to the three-year total program revenue. Each week adds the contribution to total program revenue made by the new affiliates that appear that week.

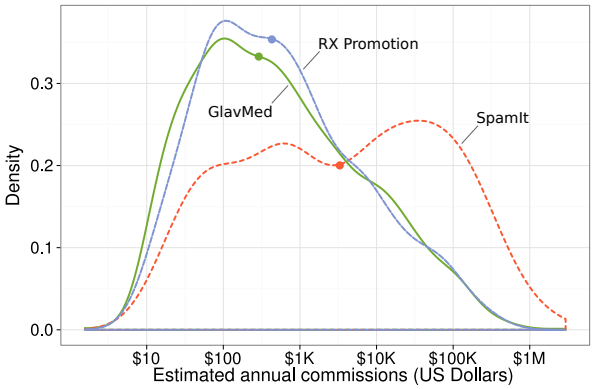


Figure 8: Distribution of affiliate commissions in each program.

ergy on the top performing affiliates. Alternatively, from an intervention perspective, undermining the activities of just a handful of affiliates would have a considerable affect on a program's bottom line: undermining the top 3–10 affiliates would impact 25–40% of program revenue.

Moreover, there is evidence that these high-revenue affiliates are not simply lucky, but represent the best-established and experienced advertisers. Figure 7 shows that it is the oldest affiliates who contribute most to weekly program revenue on an ongoing basis. For both programs, the curves show the cumulative contribution to total program revenue over time for new affiliates. For the new affiliates that appear each week, we increment a running sum with the total revenue those affiliates generate for the program—revenue generated from the moment they join until the end of the measurement period. For instance, the affiliates that generate revenue in the first week account for nearly 10% of all revenue for the entire three years of business. The dashed lines show the contributions to total revenue by affiliates that have joined on year intervals, emphasizing that the older affiliates are important for generating revenue over time. Affiliates that joined before 2008 contributed 69% GlavMed and 54% of SpamIt total program revenue as of April 2010. In contrast, affiliates that joined in 2009 and 2010 contributed less than 10% of that total.

5.2.2 Affiliate Commissions

Since only a small fraction of affiliates account for much of the business, many affiliates earn small commissions. Indeed, the median annualized affiliate commissions for GlavMed, SpamIt, and RX-Promotion are just $292, $3,320, and $428, respectively. This skew dovetails with suggestions that spam-based advertising may be a labor "lemon market" [5]. On the other hand, the most successful affiliates are able to derive substantial income through their advertising. Indeed, the top five affiliates were able to earn over $1M for themselves in a twelve-month period (and a dozen exceeded $500K).[14] Virtually all of these earnings result from sales commissions with only a minor share deriving from referral commissions (i.e., referral commissions are not a major source of income).

Figure 8 reveals a more nuanced picture of affiliate commissions. For each program, the graph shows a PDF of annualized commissions across all affiliates: the x-axis is the annualized commission earned by an affiliate, and the y-axis is the fraction of all affiliates that earned a given commission. We calculate the commission for an affiliate using the total customer sales linked to the affiliate multiplied by the commission rate of the affiliate, plus any referral commissions. Sales commission rates range from 15–45%, with 30–40% being the most common (generally high-revenue affiliates receive the highest commission rates).[15] The "dots" on the PDFs denote the median annualized commissions for that program.

For the open programs GlavMed and RX-Promotion, the majority of affiliates earn very low annualized commissions. The peaks of the PDFs range between $20–$200 a year for GlavMed, and $20–$2,000 a year for RX-Promotion. The closed program SpamIt, however, shows a bimodal distribution, with a mass of "poor" affiliates earning small commissions (mode around $500) and another mass of "rich" affiliates earning large commissions (mode around $30,000), but still with many affiliates earning over $100,000 a year.

As another perspective, on an ongoing basis the active affiliates in SpamIt, a closed program, each generate three times more revenue than active affiliates in GlavMed and RX-Promotion, both open programs. Fig-

[14]Note that Figure 8 does not involve extrapolating, but is based on taking the best four consecutive quarter's earnings for each affiliate and thus gains accuracy at the potential expense of right-censoring.

[15]Note that not all programs reward commissions uniformly over all drugs. For example, RX-Promotion typically discounts commissions by 10% on controlled drugs, so an affiliate receiving 40% on the sale of Viagra may only receive 30% on the sale of Oxycodone.

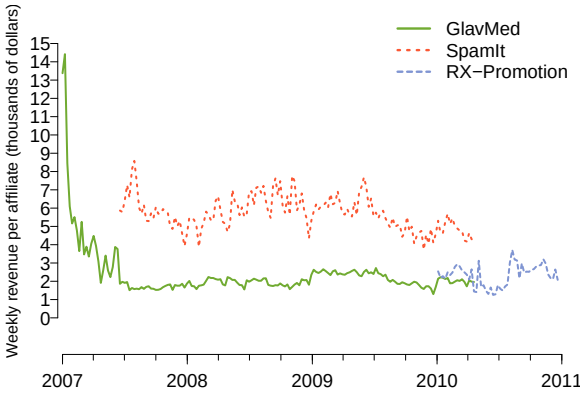


Figure 9: Average revenue per active affiliate each week.

ure 9 shows the average weekly revenue generated by active affiliates. For each week, we total the revenue generated by the affiliates that were active in attracting customers that week, and divide by the number of active affiliates. This metric is surprisingly stable over time and strongly correlates with the nature of the affiliate program. In both GlavMed and RX-Promotion, the average weekly revenue per affiliate is around $2,000. In SpamIt, though, the average weekly revenue per affiliate ranges between $5,000–$7,000. Open programs focus on increasing the total number of affiliates, but the vast majority have little impact on total revenue. Instead, by focusing on quality affiliates, the closed nature of the SpamIt program is much more effective at attracting productive affiliates and avoiding unproductive ones.

Focusing only on these most productive affiliates, we would intuitively expect them to also be the operators of the largest spamming botnets. However, even a cursory examination of the data shows that there is considerable more complexity at work. For example, while the operators of the prodigious Rustock botnet (*cosma2k*, *bird*, and *adv1*) indeed receive large commission payments (over $1.9M), botnet operators do not appear to dominate the top earners. Indeed, two of the largest botnet operators, *docent* (operator of MegaD) and *severa* (operator of Storm and Waledac) only received modest payments of $308K and $169K, respectively, for directly advertising SpamIt sites.[16]

There are a number of potential reasons for these results. First, we are only privy to sales for these particular affiliate programs and thus, if a botnet devotes much of its resources to another program, those earnings are outside our analysis. Moreover, while some botnets are largely monopolized by their owners, in many other cases the botnets are rented to provide service for third parties. For example, the second most profitable affiliate, *scorrp2*, earned close to $3M while advertising domains that we witnessed emerging from a range of botnets including MegaD, Cutwail and Xarvester. Adding to the confusion, in a number of cases botnet code is sold between parties and, thus, what some researchers may identify as a single botnet may in fact reflect multiple distinct infrastructures. Finally, we also note spamming is not the only profitable advertising vector. Indeed, the largest overall earner, *webplanet*, appears to have earned $4.6M using Web-based advertising instead. Fully unraveling the complexities of these relationships and why certain affiliates are more successful than others remains an open question.

[16] We identify botnet operators through metadata, documented more fully in the many articles in the "PharmaWars" series [10], and corroborated based on which affiliates receive money for domains known to be advertised via particular botnets.

5.3 Costs

Affiliate programs operate a complex business. As such, they have a range of costs and overheads to cover and only a fraction of their revenue translates to profit. Using a combination of transactional and metadata, we next reconstruct both direct and indirect costs for the programs. We also explore in more detail the cost structure of fulfillment (drug markup and shipping).

5.3.1 Direct Costs and Gross Margin

Direct costs are costs attributable to individual sales. While advertising is normally considered an indirect cost, affiliate programs pay for advertising as a direct cost of a sale, so we consider affiliate commissions to be a direct cost in this context. In addition, direct costs include the supplier costs for the products themselves, shipping them to customers, the fees charged by banks and credit card processors for processing customer credit card transactions, and customer refunds.

However, of these quantities only commissions are completely unambiguously encoded across all transactional data sets; RX-Promotion also includes a measure of the supplier cost and a field indicating the type of shipping (from which the shipping cost can be reverse engineered). The situation with GlavMed and SpamIt is more complex. Starting on August 8, 2008 both databases include fine-grained information about shipping and supply cost for each order. For periods before this, we are forced to extrapolate. Refunds can be calculated directly in the SpamIt and GlavMed data sets; for RX-Promotion, we infer refunds based on orders with a cancelled status. Finally, processing charges can vary among payment processors, currencies, card brands and over time. However, in examining a large number of recorded fees (found in the chatlogs) over the full period these fees range between 7–12% in practice, so as an approximation we use 10%.

Putting this data together, Table 3 itemizes the gross revenue and direct cost breakdown for GlavMed and

	GlavMed & SpamIt								RX-Promotion	
	2007		2008		2009		2010		2010	
Gross revenue	$27.3M		$60.1M		$67.7M		$18.0M		$12.8M	
Direct costs	$17.2M	(63.1%)	$42.9M	(71.4%)	$45.6M	(67.3%)	$12.1M	(67.1%)	$9.9M	(77.1%)
Commissions	$7.9M	(28.9%)	$23.0M	(38.3%)	$24.9M	(36.8%)	$6.6M	(36.7%)	$3.9M	(30.2%)
Suppliers (goods)[a]	$1.9M	(7%)	$4.3M	(7.2%)	$4.2M	(6.2%)	$1.1M	(6.1%)	$1.0M	(7.6%)
Suppliers (shipping)[b]	$3.1M	(11.4%)	$7.6M	(12.6%)	$7.8M	(11.5%)	$2.1M	(11.7%)	$1.5M	(11.5%)
Processing[c]	$2.7M	(10%)	$6.0M	(10%)	$6.8M	(10%)	$1.8M	(10%)	$1.3M	(10%)
Refunds	$1.6M	(5.9%)	$2.0M	(3.3%)	$1.9M	(2.8%)	$0.5M	(2.6%)	$1.0M	(7.8%)
Gross margin	$10.1M	(36.9%)	$17.2M	(28.6%)	$22.1M	(32.7%)	$5.9M	(32.9%)	$2.9M	(22.9%)

[a] Average supplier costs used to estimate missing supplier costs for 35% of goods.
[b] Average shipping costs used to estimate missing shipping costs for 60% of orders.
[c] Processor costs range between 7% and 11% of sales revenue.

Table 3: Gross revenue, direct costs and resulting gross margin for the GlavMed and SpamIt programs combined.

SpamIt (combined) and RX-Promotion on a yearly basis. Not surprisingly (given average affiliate commissions of 30–40%) direct costs consume the majority of revenue. Note that, due to holdback charges, the gross margin number likely overstates cash flow by around 10%, and may in fact overstate revenue as well (if holdback charges are not released). Payment processors comporting with "high risk" merchants such as these universally hold back a portion of net proceeds to handle future chargebacks and fines. From examining the logs, a 10% holdback of up to 180 days is common and, in reviewing discussions about holdbacks, the operators of GlavMed/SpamIt routinely operate under the assumption that this money may never be made available.

5.3.2 Indirect Costs and Net Revenue

Indirect costs are costs that are not generally attributable to individual sales. For online pharmacies, indirect costs are incurred for marketing (i.e., advertising the affiliate program on popular blogs and forums to attract new *affiliates*), for IT (i.e., registering domains for affiliates to use in URLs that link to storefront pages, as well as server and hosting costs), for administrative costs (i.e., staff salaries), customer service, bank fines and "lobbying". By also calculating indirect costs, we can then estimate a program's net profit—its proverbial "bottom line."

However, indirect costs are difficult to extract from transaction data since they are necessarily *indirect*. Thus, for this analysis we focus in particular on RX-Promotion for which we have highly detailed metadata comprising the raw monthly balance sheets (in spreadsheet form) for seven months of revenue. The full spreadsheet is too large to reproduce here, but we have extracted the equivalent direct costs that we calculated from transactional data in Table 3, and aggregated indirect costs in key areas. We summarize the resulting balance sheet in Table 4, reflecting seven months of revenue between March and September in 2010.

The direct costs taken from the balance sheet data are highly similar to the transactional equivalents, differing primarily due to differences in the make-up of commission tiers during this seven-month period and the greater precision available for payment processing overheads. Overall indirect costs represent almost 13% of gross, split among a range of different overheads. Note that the $157K lobbying charge is concentrated in two large payments which may be related to conflict between RX-Promotion and GlavMed/SpamIt. Overall, the net revenue for this period—the profit returned to the affiliate program owners—is just 16.3% of gross revenue. This value is not uniform from month to month, however. For example, during the period when processing for controlled drugs was lost, RX-Promotion simultaneously lost revenue, incurred large fines, and had to pay greater average commissions (since the commissions for controlled drugs were discounted 10%) leading to a net

	RX-Promotion March – September 2010	
Gross revenue	$7.8M	
Direct costs	$5.5M	(70.8%)
Commissions	$3M	(38.1%)
Suppliers[a]	$1.4M	(17.6%)
Processing	$1M	(13.2%)
Other direct	$148.3K	(1.9%)
Indirect costs	$1004K	(12.8%)
Administrative	$197K	(2.5%)
Customer service	$124K	(1.6%)
Fines	$107K	(1.4%)
IT expenses	$202K	(2.6%)
Domains	$114K	(1.5%)
Servers, hosting	$66K	(0.8%)
Selling expenses	$315K	(4%)
Marketing	$105K	(1.3%)
Lobbying	$157K	(2%)
Other indirect	$134K	(1.7%)
Net revenue	$1.3M	(16.3%)

[a] Costs of goods and shipping are combined.

Table 4: Balance sheet for RX-Promotion detailing indirect costs.

loss for at least one month. By contrast, during the very best month (September) net revenue exceeds 30%.

We do not have equivalent indirect cost data for GlavMed or SpamIt, but we are able to infer a subset of these overheads. The operators used a special affiliate (`affiliate_id` value 20) to manage the working capital of each. The Affiliate 20 account received referral commissions from all affiliates who did not have a referring affiliate designated explicitly. During the measurement period, Affiliate 20 earned $2.7M. Operating expenditures, as well as some affiliate payouts, were deducted from this account.

Starting May 2009, the `comment` field of each payout began including a short description of the payment. A payment for a banner advertisement (recruiting affiliates), for example, would be listed as described as "banner GM - gofuckbiz.com". Although free-form, the comment text typically used a small number of phrases. Using a manually generated list of regular expressions, we identified several indirect costs during the period from May 2009 to April 2010. These costs include marketing ($153K, 0.2% of revenue), domain purchasing ($511K, 0.8% of revenue) and servers/hosting ($247K, 0.4% of revenue). Interestingly, it appears that marketing and servers/hosting are similar costs between the two programs (suggesting they are largely fixed costs) but domain purchasing appears to track revenue (presumably since greater advertising volume requires more domain turnover due to blacklisting).

Finally, we also have anecdotal data in the form of chat logs between the lead operator and the owner of GlavMed/SpamIt. These logs state that overall net revenue fluctuated between 10% and 20%, agreeing structurally with the RX-Promotion data.

Thus, we believe that 10–20% is likely to reflect a typical net revenue for successful pharmaceutical programs. While this is smaller on an earnings-per-sale basis than the commissions awarded to individual affiliates, it is a more profitable enterprise when the affiliate program is successful. For example, the largest SpamIt affiliate might make $2M in a year, but in that same year the program itself is likely to clear over $10M in profit.

5.3.3 Markup

After commissions, supply costs for the programs are one of the largest expenses. Using the categories from Figure 2, ED contains by far the most popular products purchased, and also has the highest markups of more than 15 to 20 *times* the supply cost. The average markup of Viagra in GlavMed and SpamIt, for instance, translates to a customer price 25 times cost. Markups in the Abuse and Chronic categories are considerably smaller, ranging between 5–8 times supply cost. Interestingly, the shipping cost is a loss leader for GlavMed/SpamIt since they charge a flat fee per order (orders with more than one item result in supplier shipping costs higher than collected shipping fees) and offer free shipping for orders over $200. In fact, for the orders for which we have fine-grained product and shipping cost data, the supplier costs of delivering the drugs (8.5M) actually exceeded the costs of the drugs delivered.

5.4 Payment Processing

Finally, affiliate programs must arrange for reliable processing of customer payments. In a sense, obtaining reliable payment processing services may be the most important function of the affiliate program, since it is the only mechanism by which all other efforts can be monetized. Previously, our group identified that a small number of banks were critical to virtually all online pharmaceutical sales [14]. However, the means by which those banks were accessed has never been well documented.

In fact, in the "high-risk" payment market, merchant processing is frequently handled by independent Payment Service Providers (PSPs) who manage the relationships with acquiring banks and provide Web-based payment gateway services to clients.[17] While users of these services may have a contractual relationship with the bank, in other cases PSPs may "front" their own merchant accounts on behalf of their clients (a form of identity laundering called "factoring" and typically disallowed by card association rules). Merchants in turn can mitigate some of their own risk by working with multiple providers; this strategy not only provides redundancy, but each provider may place limits on transaction volumes (e.g., to fit within the underwriting risk limits on their overall merchant portfolio) and may have different services they are willing to offer (e.g., MC, Visa, Amex, eCheck, etc.) for different product categories (e.g., herbal vs. prescription vs. controlled drugs).

In the case of RX-Promotion the affiliate program enjoyed a partnership with a large ISO/PSP and thus this entity handled virtually all of their processing needs. GlavMed and SpamIt, by contrast, did not work with any single provider, but no less that twenty-one distinct providers over the lifetime of our data sets. However, these providers differ considerably in what services they are used for, the volume of transactions they are able to handle and how long-lived they are. In fact, almost half of these providers are never used to process significant transaction volumes (mostly likely due to risk controls).

Illustrating this point, Figure 10 graphs the transaction volume of GlavMed/SpamIt handled by different payment service providers over time. The *y*-axis identifies

[17] We use the term "payment service provider" here in a generic sense, and the organizations involved may be some combination of proper PSPs, account brokers, merchant servicers, ISO/MSPs with third-party servicers, etc.

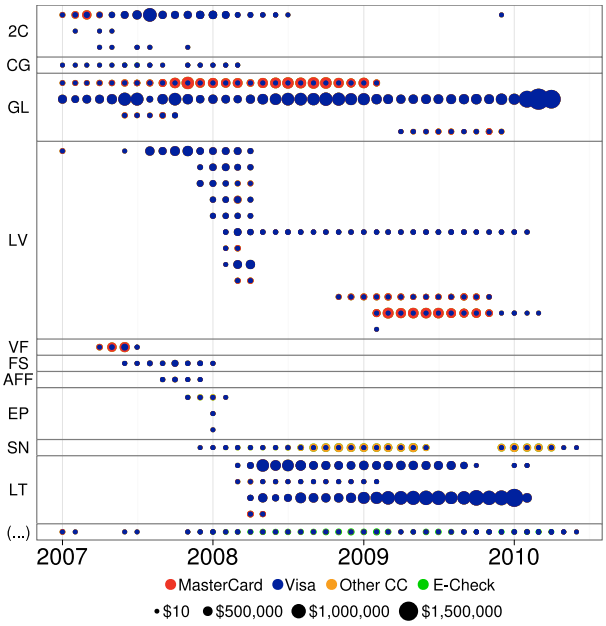


Figure 10: Payment transactions over time by payment service provider. The colored volume of each circle corresponds to the transaction volume in a month for a particular terminal (color indicating payment method), with terminals grouped by providers.

the top nine providers (using a designator taken directly from the database or an abbreviation thereof) while the remaining providers are aggregated together under the ellipsis. Each circle in the graph represents the number of transactions processed via a particular *terminal* in a month, with terminals belonging to a particular provider grouped together based on time of first use.[18] In any given circle, the color red indicates MasterCard transactions, blue is for Visa, yellow for other credit cards (primarily Amex), and green for eCheck.

There are a number of striking observations one can draw from this figure. First is the clear dominance of Visa processing. Aggregating across both GlavMed and SpamIt, Visa transactions represent almost 67% of all revenue, followed by MasterCard with 23% and American Express with 6% (with the remainder concentrated in eCheck transactions through the ACH system). While part of this discrepancy is likely due to demand—Visa is the most popular payment card brand—this difference also reflects a supply issue as well. For reasons not entirely clear, it has traditionally been far easier for online pharmaceutical programs to obtain payment processing services for Visa than for MasterCard or Amex. Indeed, we find that during periods in which MasterCard processing was *available*, Visa/MasterCard revenue percentages stabilized at around 63%/30%, respectively, for both GlavMed and SpamIt.

Second, a relatively small number of payment service providers dominate the transaction volume (in particular GL, LT and LV). Together these three providers are responsible for 84% of all revenue for GlavMed and SpamIt. Many of the other providers are active for very short lifetimes, and with very low volumes, before they are either abandoned or, more typically, they are unwilling to continue business with the program operators.

Finally, there are also clear patterns indicative of problems with particular providers over time. For example, for each terminal a sudden drop in volume and rise in declines (not shown) is typically a precursor to that terminal being abandoned. Some of these cases clearly reflect changes in long-term business relationships: in March of 2008, for instance, there is a clear transition moving the largest volume of Visa processing between LV and LT; similarly, American Express processing moves from AFF to SN during the same period. In the last five months of 2010 it appears that GlavMed/SpamIt experienced significant setbacks in processing capability, with LT processing only minor volumes (forcing them to push a higher volume of transactions through GL). These findings provide additional support and context for our previous findings that the financial aspect of the counterfeit pharmaceutical ecosystem is among the most fragile components [14].

6 Conclusion

This paper provides an unprecedented view inside the economics of modern pharmaceutical affiliate programs: an enterprise that ultimately capitalizes a wide array of infrastructure services including botnets, malware, bullet-proof hosting and so on. Among the results of this work, we have shown that the customer market is large and far from fully tapped, with repeat orders playing a key role in mature programs. We have also seen that a small number of big affiliates can dominate the revenue equation and that disrupting these particular affiliates would have disproportionate damage on the whole program. Finally, even very large programs such as GlavMed/SpamIt depend on a handful of payment service providers to reliably monetize their activities, reinforcing the observation that financial services are a "weak point" in the value chain. Surprisingly, while affiliate programs can drive substantial sales, their costs are significant and ultimately net revenues are modest, typically under just 20% of sales. This finding again suggests that such organizations are fragile to economic disruptions of even a modest scale.

[18] A terminal is effectively an interface point for sending payment transactions, corresponding to a particular merchant account. Note that while some terminals are for general purpose use, others service a particular function such as providing a compatible base currency (e.g., the terminal named "lt-euro-visa" provides European Visa transactions) or handling rebills (e.g., "gl-rebill-m").

Acknowledgments

We would like to thank the various anonymous providers of our data sets, without which there would have been no paper. We have also benefited heavily from the many members of the cyber-investigations community who have provided us valuable insight as we have tried to map data onto meaning. Closer to home, we would like to thank Erin Kenneally for her ongoing legal guidance and ethical oversight, as well as the technical support of Brian Kantor and Cindy Moore who have managed our systems and storage needs.

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).

References

[1] D. S. Anderson, C. Fleizach, S. Savage, and G. M. Voelker. Spamscatter: Characterizing Internet Scam Hosting Infrastructure. In *Proc. of 16th USENIX Security*, 2007.

[2] Behind Online Pharma. From Mumbai to Riga to New York: Our Investigative Class Follows the Trail of Illegal Pharma. http://behindonlinepharma.com, 2009.

[3] J. Caballero, C. Grier, C. Kreibich, and V. Paxson. Measuring Pay-per-Install: The Commoditization of Malware Distribution. In *Proc. of 20th USENIX Security*, 2011.

[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. Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy. In *Proc. of 8th WEIS*, 2009.

[6] J. P. John, A. Moshchuk, S. D. Gribble, and A. Krishnamurthy. Studying Spamming Botnets Using Botlab. In *Proc. of 6th NSDI*, 2009.

[7] J. P. John, F. Yu, Y. Xie, A. Krishnamurthy, and M. Abadi. deSEO: Combating Search-Result Poisoning. In *Proc. of 20th USENIX Security*, 2011.

[8] 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.

[9] C. Kanich, N. Weaver, D. McCoy, T. Halvorson, C. Kreibich, K. Levchenko, V. Paxson, G. M. Voelker, and S. Savage. Show Me the Money: Characterizing Spam-advertised Revenue. In *Proc. of 20th USENIX Security*, 2011.

[10] B. Krebs. SpamIt, Glavmed Pharmacy Networks Exposed. Krebs on Security Blog, http://www.krebsonsecurity.com/category/pharma-wars/, 2011.

[11] LegitScript and KnujOn. No Prescription Required: Bing.com Prescription Drug Ads. http://www.legitscript.com/download/BingRxReport.pdf, 2009.

[12] LegitScript and KnujOn. Yahoo! Internet Pharmacy Advertisements. http://www.legitscript.com/download/YahooRxAnalysis.pdf, 2009.

[13] N. Leontiadis, T. Moore, and N. Christin. Measuring and Analyzing Search-Redirection Attacks in the Illicit Online Prescription Drug Trade. In *Proc. 20th USENIX Security*, 2011.

[14] K. Levchenko, N. Chachra, B. Enright, M. Felegyhazi, C. Grier, T. Halvorson, C. Kanich, C. Kreibich, H. Liu, D. McCoy, A. Pitsillidis, N. Weaver, V. Paxson, G. M. Voelker, and S. Savage. Click Trajectories: End-to-End Analysis of the Spam Value Chain. In *Proc. of 32nd IEEE Security and Privacy*, 2011.

[15] H. Liu, K. Levchenko, M. Félegyházi, C. Kreibich, G. Maier, G. M. Voelker, and S. Savage. On the Effects of Registrar-level Intervention. In *Proc. of 4th USENIX LEET*, 2011.

[16] B. S. McWilliams. *Spam Kings: The Real Story Behind the High-Rolling Hucksters Pushing Porn, Pills and @*#?% Enlargements*. O'Reilly Media, Sept. 2004.

[17] A. Ramachandran and N. Feamster. Understanding the Network-Level Behavior of Spammers. In *Proceedings of ACM SIGCOMM*, Pisa, Italy, Sept. 2006.

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

[19] Y. Shin, M. Gupta, and S. Myers. The Nuts and Bolts of a Forum Spam Automator. In *Proc. of 4th USENIX LEET*, 2011.

[20] B. Stone-Gross, R. Abman, R. Kemmerer, C. Kruegel, D. Steigerwald, and G. Vigna. The Underground Economy of Fake Antivirus Software. In *Proc. of 10th WEIS*, 2011.

[21] Symantec. MessageLabs June 2010 Intelligence Report. http://www.symanteccloud.com/mlireport/MLI_2010_06_June_FINAL.pdf.

[22] K. Thomas, C. Grier, V. Paxson, and D. Song. Suspended Accounts In Retrospect: An Analysis of Twitter Spam. In *Proc. of 11th IMC*, 2011.

[23] D. Wang, S. Savage, and G. M. Voelker. Cloak and Dagger: Dynamics of Web Search Cloaking. In *Proc. of 18th CCS*, 2011.

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

[25] G. Wondracek, T. Holz, C. Platzer, E. Kirda, and C. Kruegel. Is the Internet for Porn? An Insight into the Online Adult Industry. In *Proc. of 9th WEIS*, 2010.