An Inquiry into the Nature and Causes of the Wealth of Internet Miscreants*

Jason Franklin
Carnegie Mellon University
jfrankli@cs.cmu.edu

Vern Paxson
ICSI
vern@icsi.berkeley.edu

Adrian Perrig
Cylab/CMU
perrig@cmu.edu

Stefan Savage
UC San Diego
savage@cs.ucsd.edu

ABSTRACT

This paper studies an active underground economy which specializes in the commoditization of activities such as credit card fraud, identity theft, spamming, phishing, online credential theft, and the sale of compromised hosts. Using a seven month trace of logs collected from an active underground market operating on public Internet chat networks, we measure how the shift from "hacking for fun" to "hacking for profit" has given birth to a societal substrate mature enough to steal wealth into the millions of dollars in less than one year.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: ABUSE AND CRIME INVOLVING COMPUTERS

General Terms

Security

Keywords

Measurement, Underground Markets, E-crime

*This research was supported in part by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office, and grants CNS-0537246, CCF-0424422, CNS-0433668, and NSF-0433702 from the National Science Foundation. Jason Franklin performed this research while on appointment as a U.S. Department of Homeland Security (DHS) Fellow under the DHS Scholarship and Fellowship Program, a program administered by the Oak Ridge Institute for Science and Education (ORISE) for DHS through an interagency agreement with the U.S Department of Energy (DOE). The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of ARO, CMU, NSF, DHS, DOE, ORISE or the U.S. Government or any of its agencies.

CCS'07, October 29–November 2, 2007, Alexandria, Virginia, USA.

1. INTRODUCTION

Computer security is a field that lives in co-dependence with an adversary. The motivation for security research is ever to stymie the goals of some hypothetical miscreant determined to violate one of our security policies. Typically, we abstract away their motivations and consider the adversary solely in terms of their capabilities. There is good reason for this since the threat model for any security mechanism is generally driven entirely by the adversary's abilities. Moreover, reasoning about any individual's state of mind, let alone predicting their behavior, is inherently prone to error. That said, the nature of Internet-based threats has changed over the last decade in ways that make it compelling to attempt a better understanding of today's adversaries and the mechanisms by which they are driven.

First and foremost among these changes is the widespread observation that Internet-based criminal activity has been transformed from a reputation economy (i.e., receiving "street cred" for defacing Web sites or authoring viruses) to a cash economy (e.g., via SPAM, phishing, DDoS extortion, etc). Indeed, even legal activities such as vulnerability research has been pulled by the gravity of a cash economy and today new vulnerabilities are routinely bought and sold by public companies and underground organizations alike [12]. Thus, there is a large fraction of Internet-based crime that is now fundamentally profit driven and can be modeled roughly as rational behavior. Second, and more importantly, the nature of this activity has expanded and evolved to the point where it exceeds the capacity of a closed group. In fact, there is an active and diverse on-line market economy that trades in illicit digital goods and services in the support of criminal activities. Thus, while any individual miscreant may be difficult to analyze, analyzing the overall market behavior and the forces acting on it is far more feasible.

This paper is a first exploration into measuring and analyzing this market economy. Using a dataset collected over 7 months and comprising over 13 million messages, we document a large illicit market, categorize the participants and explore the goods and services offered. It is our belief that better understanding the underground market will offer insight into measuring threats, how to prioritize defenses and, ultimately, may identify vulnerabilities in the underground economy itself.

The paper is organized as follows. Section 2 provides an overview of the market being studied. Section 3 is an analysis of relevant issues including market significance, participation, and services. Section 4 measures the advertisements seen in the market and provides price data. Section 5 discusses applications of our measurements and countermeasures to disrupt the market. Sections 6 and 7 present related work and our conclusions.

2. MARKET OVERVIEW

The market studied in this paper is a public channel commonly found on Internet Relay Chat (IRC) networks. It provides buyers and sellers with a meeting place to buy, sell, and trade goods and services in support of activities such as credit card fraud, identity theft, spamming, phishing, online credential theft, and the sale of compromised hosts, among others.

2.1 IRC Background

Internet Relay Chat (IRC) is a standard protocol for real-time message exchange over the Internet [13]. IRC employs a client-server model where clients connect to an IRC server which may peer with other servers to form an IRC network.

To connect to an IRC network, an IRC client first looks up the address of a server belonging to the network then connects to the network by way of the server. After connecting, the client identifies itself with an IRC nickname (nick) which can be registered by assigning a password. To begin communicating, a client typically queries the network for the list of all named communication areas known as channels.

After joining a named channel, a client can send both public (one-to-many) and private (one-to-one) messages. Public messages are broadcast to all clients connected to the channel. Private messages are transmitted from the source client to the destination client without being displayed in the channel. Private messages pass through any intermediate IRC servers between the source and destination, but are not available to the other clients connected to the channel including the channel administrators, called channel operators.

2.2 Data Collection

Our dataset is comprised of 2.4GBs of Internet Relay Chat (IRC) logs archived over a 7 month period ranging from January to August of 2006. The logs were collected by connecting to a particular channel on different IRC networks and logging all subsequent public messages. Each log is of the format: (timestamp, IRC server IP address, source identifier, channel name, message). The dataset contains over 13 million messages from a total of more than one hundred thousand distinct nicks.

The IRC channels monitored were simultaneously active on a number of independent IRC networks. Each network provides a separate channel which may include over three hundred participants at any time. While the channels on each network are separate, the predominance of certain types of common activities establish uniformity across networks and create a market.

2.3 Market Administration

Channel administrators are responsible for the well-being of the market including maintaining a list of verified participants, enforcing client identification policies, and running an automated channel service bot.

Verified Participants. A culture of dishonesty and distrust pervades the market making it necessary to differentiate trustworthy individuals from dishonest "rippers," individuals who conduct fraudulent transactions. To facilitate honest transactions, channel administrators provide a participant verification service. After a nick demonstrates their trustworthiness, they are given a special designation, $+v$ (the IRC 'voice' attribute), as a seal of approval from the channel's administrators.

Channel administrators continuously remind buyers and sellers to only undertake transactions with other verified participants. Channel participants look for the $+v$ designation to determine the level of care to take when dealing with a particular nick. Many participants only undertake transactions with other verified nicks or require unverified participants to complete their end of the transaction first to ensure the unverified participant upholds their end of the deal.

Client Identification. Each line of data in the corpus contains a source identifier for the client who sent the message to the channel. The source identifier contains three fields: an IRC nick, a client username or Ident [10] response, and a host identifier such as an IP address or hostname. The nick and host identifier fields are used in the market for client identification. Upon connecting to an IRC server, a client's IP address may be checked against a local, block list used to prevent access from unruly IPs or to prevent client access from anonymization services. A client's IRC nick may be checked against a local database of previously registered names. If the client's nick was previously registered, a password is required to use the nick. Otherwise, the client may proceed as an unregistered user or register their nick by assigning a password. Registration is a necessary, first step for clients who wish to build business relationships or sometimes even post to a channel. Finally, the market administrators maintain a list of registered nicks which belong to verified participants.

Channel Services. Most networks include one or more automated channel service bots which provide a myriad of interactive services including credit card limit lookups, credit card validation code (CVV2) lookups, listing IP addresses of open proxies, returning e-merchants who perform limited credit card authorization checks, and tracking the time a nick was last active.

2.4 Market Activity

The majority of the public messages in the market can be broadly categorized into two types: advertisements and sensitive data.

Advertisements. The most common behavior in the market is the posting of want and sales ads for illicit digital goods and services. Goods range from compromised machines to mass email lists for spamming. Services range from electronically transferring funds out of bank accounts to spamming and phishing for hire. Table 1 includes actual ads seen in the market and their meanings.

The goods and services advertised are sold to miscreants who perform various forms of e-crime including financial fraud, phishing, and spamming. For example, a miscreant, intent on phishing, can enter the market and buy the goods necessary to launch a targeted phishing campaign: targeted email addresses derived from web crawling or compromised databases, mailers installed on compromised hosts or web forms vulnerable to email injection attacks [1], compromised machines to host the phishing "scam" page, and software which promises to bypass spam filters. Similarly, a miscreant, intent on committing financial fraud, can enter the market and purchase credentials such as bank logins and passwords, PayPal accounts, credit cards, and social security numbers (SSNs). After purchasing credentials, the fraudster may employ the services of a "cashier," a miscreant who specializes in the conversion of financial credentials into funds. To perform their task, the cashiers may work with a "confirmer," a miscreant who poses as the sender in a money transfer using a stolen account. After each miscreant performs their task, the fraudster's transaction is complete and the supporting miscreants typically accept their payment through an online currency such as E-Gold or an offline source such as Western Union money transfer.

Sensitive Data. The second most common behavior in the market is pasting sensitive data to the channel. For example, it is common to see miscreants post sensitive data such as the following credit card and identity information:

Advertisement	Classification Label(s)
i have boa wells and barclays bank logins....	Bank Login Sale Ad
have hacked hosts, mail lists, php mailer send to all inbox	Hacked Host Sale Ad, Mailing List Sale Ad, Mailer Sale Ad
i need 1 mastercard i give 1 linux hacked root	Credit Card Want Ad, Hacked Host Sale Ad
i have verified paypal accounts with good balance...and i can cashout paypals	PayPal Sale Ad, Cashier Service Ad

Table 1: Advertisements with labels used for classification.

```
Name: Phil Phished
Address: 100 Scammed Lane, Pittsburgh, PA
Phone: 555-687-5309
Card Number: 4123 4567 8901 2345
Exp: 10/09    CVV: 123
SSN: 123-45-6789
```

Sensitive data posted to the channel may or may not include sufficient information to make it immediately useful to other channel members. In the credit card information example, other channel members could begin using Phil's card or steal his identity. Other times, sensitive data may be posted to the channel as a way to demonstrate the existence of a valuable commodity such as access to a financial account without giving the commodity away. For example, miscreants post partial account numbers along with their balances as a form of sales ad.

```
CHECKING 123-456-XXXX $51,337.31
SAVINGS 987-654-XXXX $75,299.64
```

Sensitive data may be either explicitly labeled as in the previous examples or posted without a label. When explicitly flagged, a miscreant intentionally appends a label to the data before posting to the channel. This label helps to identify the data type and disambiguate fields. However not all sensitive data is labeled. Often miscreants simply paste sensitive data under the assumption that fields such as names, addresses, and phone numbers are implicitly recognizable. Since relying on labels would limit the extent to which data could be measured, the measurements in this paper use pattern matches for structured data such as credit cards and social security numbers and random sampling in conjunction with manual labeling for free form data such as names, addresses, and usernames and passwords.

2.5 Measurement Methodology

This paper contains three classes of measurements: manual, syntactic, and semantic. The primary differences between classes are the techniques used and their level of accuracy.

Manual Measurements. We hand labeled a 3,789 line dataset selected uniformly at random from the corpus with several dozen labels describing the goods and services advertised and sensitive data in each message. Labels describing ads include the good or service being advertised and the type of advertisement (want or sale). Labels describing sensitive data signify the data type (e.g., credit card number, CVV2, SSN, etc.). Table 1 includes real ads with their corresponding category labels. Throughout the remainder of the paper, references to labeled data or the labeled dataset are meant to denote this manually labeled data.

Syntactic Measurements. Syntactic measurements use pattern matches in the form of regular expressions and achieve a high degree of accuracy. When necessary, both matches and mismatches are measured. Other measurements which fall into this category include the use of the Luhn algorithm to verify credit card numbers, IP address lookups on DNS blacklists, and social security number lookups in a Social Security Administration database.

Semantic Measurements. Semantic measurements make use of supervised machine learning techniques to classify text into more than sixty categories with associated meanings. To automatically classify ads such as those in Table 1, we use statistical machine learning classifiers to label each line with an associated meaning. In particular, we employ linear support vector machines (SVMs) with bag-of-words feature vectors, term frequency-inverse document frequency (TFIDF) feature representation, and an L2-norm as implemented in the SVM^{light} package [9]. Similar approaches have been used in the past for accurate and scalable classification of large text corpora [5, 20].

We split the labeled dataset chronologically, the first 70% was used as a training set and the remaining 30% as a test set. We trained a binary SVM classifier for each of our categories. We performed offline classification of the 13 million unlabeled messages to identify ads throughout the monitored period. Measurements made with the SVM classifiers contain both false positives and false negatives and are accompanied by performance statistics from the test set.

2.6 Complexities and Limitations

Several limitations and complexities underlie the measurements and analysis in this paper.

Market Visibility. The dataset used in this paper does not contain private messages between participants. Private messages contain the majority of transaction details. The measurements in this paper include public messages and ads sent to every client in the channel.

Assertions versus Intentions. Assertions do not necessarily represent the underlying intentions of a market participant. For example, a seller may advertise social security numbers for sale with the intention of tricking unsuspecting buyers into paying before receiving SSNs. The measurements in this paper use aggregation and statistical analysis to minimize the effect of dishonest advertising.

Monitored Individuals Biasing Analysis. Individuals who know they are being monitored may change their behavior resulting in skewed measurements. The anonymity provided by proxies and the market's focus on illegal activities makes such behavior unlikely.

3. MARKET ANALYSIS

We begin our analysis of the underground market by asking a necessary preliminary question: "Is the market significant?" To answer this question, we measure the extent to which the market enables identity theft, credit card fraud, and other illicit activities. Next, we build a profile of the market's members by measuring market participation including activity levels, participant lifetimes, verified status; and correlating participant's IPs with known exploited IPs, proxies, and IPs which send spam. Finally, we analyze the services provided by the market's administrators and discuss the incentives behind operating an underground market.

3.1 Sensitive Data and Market Significance

In order to establish the significance of the market being studied, we present measurements of the sensitive data observed in the open market. We believe sensitive data is posted to the channel for two primary reasons: 1) sellers providing samples of useful data such as credit card data to build credibility or demonstrate that they possess valid data and 2) participants submitting sensitive data in queries to the channel services bot.

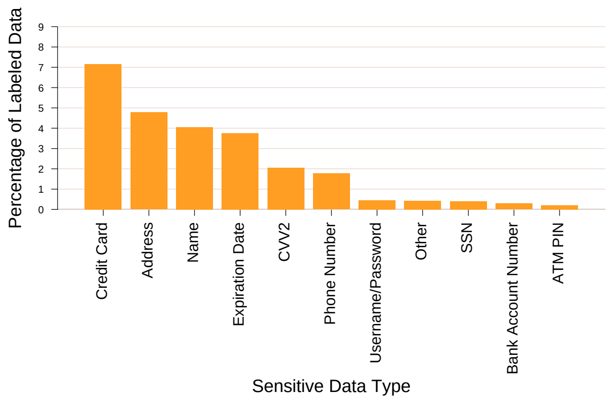


Figure 1: Sensitive data distribution in labeled dataset.

***Sensitive Data*: Measurement Methodology.** To determine the extent to which posting sensitive data pervades the market, we count the number of messages in the manually-labeled dataset which contain sensitive data including credit card numbers and expiration dates, addresses, names, Card Verification Values (CVV2s), phone numbers, usernames and passwords, mother's maiden names, answers to challenge questions, SSNs, bank account numbers, ATM pins, driver's license numbers, and dates of birth. Since we are establishing an upper bound on the levels of sensitive data, we do not remove repeated data nor do we verify the validity of the sensitive data found. Subsequent measurements address the issues of data repetition and validation.

***Sensitive Data*: Measurement Results.** The percentage of messages containing various types of sensitive data is shown in Figure 1. These measurements show that by randomly sampling from the 13 million line corpus a significant amount of sensitive data can be found. Furthermore, these measurements suggest that the market is awash in freely available data of all types. To understand the magnitude of the sensitive data available, we further measure the highest percentage sensitive data, credit card numbers, and two important data types: financial account data and SSNs.

3.1.1 *Credit Card Data*

***Credit Card Data*: Measurement Methodology.** We identify potential credit card numbers by pattern matching numbers which appear to be properly formatted Visa, Mastercard, American Express, or Discover cards. To maximize the number of cards identified, we use syntactic matches rather than relying on miscreants to explicitly flag their posted data. We remove repeated cards and check that each unique card number has a valid Luhn digit [11]. The Luhn digit is a checksum value which guards against simple errors in transmission. While passing the Luhn check is a necessary condition for card validity, it does not guarantee that the card number has been issued, is active, or has available credit at the time of posting.

***Credit Card Data*: Measurement Results.** Including repeats, we found a total of 974,951 credit card numbers in the corpus. This represents 7.4% of the total logs which is close to the 7.15% estimate established in Section 3.1. Eliminating duplicate values, there were a total of 100,490 unique credit card numbers. Other card numbers are present in the corpus, but their representations include text, delimiters, or other separators which resist simple pattern matches. Hence, the number of cards found can be considered a conservative estimate. The results of our measurements are summarized in Table 2.

Card Type	Valid Luhn Digit	Invalid Luhn Digit
Visa	53,321	6,540
Mastercard	26,581	6,486
American Express	5,405	265
Discover Card	1,836	56
Total	87,143	13,347

Table 2: Credit card statistics.

To correlate our data with another source, we look up a small sample of the credit cards with valid Luhn digits in TrustedID's StolenIDSearch database of 2,484,411 numbers. TrustedID states that they receive information, "by looking in places where fraudsters typically trade or store this kind of information." StolenIDSearch provides a query interface for consumers to check if their identity or credit information may be compromised. Of the 181 cards numbers we queried, 51% were in the TrustedID database as of August 2007. The high percentage of matches may be the result of TrustedID monitoring the same servers we monitored. Alternatively, the card numbers may be available in multiple locations.

To understand the possible origins of the credit card data, we manually survey the data and the flags miscreants use to identity sensitive data posted to the channel. We found over 1,300 flags which start with the prefix, "AOL". We believe this prefix is meant to designate the Internet service provider America Online and is used to flag data derived from AOL subscribers. In addition, we found tens of thousands of instances of shipping instructions embedded with delimited data which appears to be extracted from a formatted file or database containing e-merchant order information.

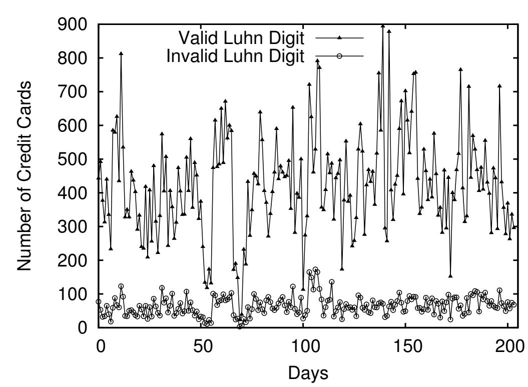
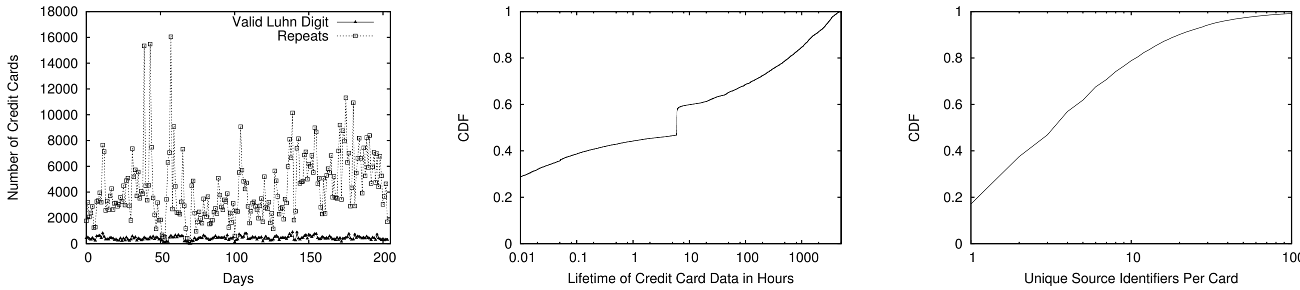


Figure 2: Credit card arrivals.

***Credit Card Arrivals*: Measurement Methodology.** To establish the number of credit cards in the channel, we measure the rate at which new data enters the channel and the rate at which previously seen cards are repeated. Repetition is typically the result of channel participants providing the same data sample multiple times, or card numbers being repeated in requests to and responses from the channel services bot.

***Credit Card Arrivals*: Measurement Results.** Figure 2 shows the arrival rates of potentially valid cards which pass the Luhn check and invalid cards which fail the Luhn check. Valid cards arrive with an average rate of 402 cards per day or close to 17 cards per hour with a standard deviation of 145 cards per day. Invalid cards arrive at an average rate of 88 cards per day. The arrival of valid card numbers at a steady rate for over 200 straight days seems to imply that miscreants either continuously collect card data through



(a) New and Repeated Card Arrivals (b) Credit Card Lifetimes (c) Unique Source IDs Per Card

Figure 3: Credit card number repetitions, lifetimes, and sources per card.

activities such as phishing or compromising merchant databases, or that miscreants possess large numbers of stolen cards. The regular arrival of invalid cards suggests that some novice miscreants lack sufficient knowledge or sophistication to use one of the many publicly available programs which generate card numbers with valid Luhn digits.

***Credit Card Repetition, Lifetime, and Sources*: Measurement Methodology.** To better understand the card data seen in the market, we measure the arrival rate of repeats, the lifetime of a card, defined as the time between the first and last post, and the number of sources which post each card. The lifetime and source measurements include cards with both valid and invalid Luhn digits. For the source measurement, we use the full source identifier including the IRC nick, username or Ident field, and hostname as the atom of client identification.

***Credit Card Repetition, Lifetime, and Sources*: Measurement Results.** Figure 3(a) shows repeated cards arriving over an order of magnitude faster than cards with valid Luhn digits at an average rate of 4,272 cards per day. The majority of cards are repeated fewer than 4 times and 95% of cards are repeated fewer than 34 times. Figure 3(b) shows that over 40% of all card numbers are seen within a half-hour period and the majority of cards are exposed for six hours or less. Figure 3(c) shows the number of sources per card. Around 17% of cards are posted by a single source (non-repeats) and the majority of cards are posted by 4 or fewer sources. The limited number of repetitions per card, the limited lifetime of most cards, and the small number of sources which post each card suggests that repeating the same data sample over and over is of limited use. It is possible that once pasted, the entire available credit limit is quickly spent or the card is removed from service by fraud prevention services monitoring the channel or monitoring card activity.

***Bank Identification Number*: Measurement Methodology.** For each unique credit card seen in the channel, we look up the bank identification number (BIN) information to ascertain the country of the issuing bank. The first six digits of a credit card, called a BIN or Issuer Identification Number (IIN), uniquely identify the country of the issuing bank, bank or organization name, funding type (Credit, Debit, or Prepaid), and card type (e.g., Classic, Gold, etc.). American Express and Discover cards do not include BIN numbers because, unlike Visas and Mastercards, they are not distributed by networks of banks but by individual companies.

The official BIN number database is not available to the public. We use a BIN list containing information for 52,492 issuing banks of Visas and Mastercards which we acquired as part of the source code of a channel service bot. We crosscheck our BIN list by looking up a small percentage (0.1%) of the BINs in a BIN database[1], currently being created as part of a community effort to publicize BIN information. We were unable to look up every BIN from the underground list in the public database since it limits the number of BIN lookups from an unique IP address to around 10 a day. When performing validation of our BIN list, the country and bank names in the public database exactly matched the underground data.

***Bank Identification Number*: Measurement Results.** To assess the extent to which credit card data from around the world finds its way into the market, we look up the country of the issuing bank of each unique Visa and Mastercard with valid Luhn digits. Of 11,649 unique BINs, 2,998 BINs representing 7.3% of Visa and 13.9% of Mastercards are not found in the BIN list. The results of our measurements are presented in Figure 4.

As one might expect of a market with a stated “English Only” policy, the majority of cards were from issuing banks in the United States (62,142) and the United Kingdom (3,977). Other countries with greater than 200 occurrences include Canada, Brazil, Australia, France, Germany, and Malaysia. While the country of origin of the issuing bank is not always the country where the card is currently being used nor the country where the data was compromised, the number of countries represented in the data suggests that the market has global data sources and that the market’s participants may be dispersed around the world.

Further evidence that the market is international can be found in the details of ads from the participants. Ads often carry restrictions on the type of data wanted or being offered or the type of buyer required. Examples include buyers placing thousands of requests for cards from Japan, Italy, India, and Pakistan and sellers whose ads include warnings such as “No nigerians or romanians!” and more colorfully worded restrictions.

3.1.2 Financial Data

In addition to credit card data, other financial data seen in the channel includes checking and savings account numbers and balances. Miscreants often post text which they purport to be copied directly from a financial account access webpages and tout screen captures of account webpages to attest to their ability to access an account with a particular balance.

***Financial Data*: Measurement Methodology.** To quantify the dollar value of the financial data posted, we sum the checking, savings, mortgage, and balance figures. We add each unique dollar amount only once to prevent double counting of balances, even across categories. While pasting financial account balances is trivial to fake and difficult to validate, the practice is used by honest

[1]http://www.bindatabase.com

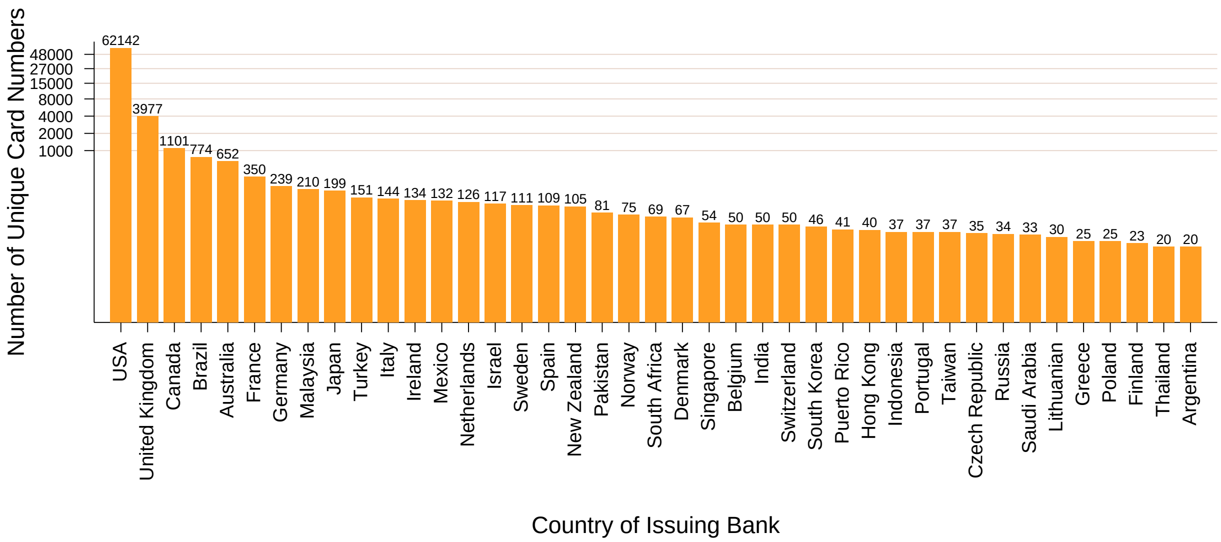


Figure 4: Issuing bank country distribution for Visas and Mastercards.

sellers to advertise actual accounts for sale. We are unable to verify the percentage of posts which are valid.

***Financial Data*: Measurement Results.** The results of our measurements are presented in Table 3.

Account Type	Total Balance
Balance	$18,653,081.08
Checking	$17,068,914.96
Mortgage	$15,892,885.37
Savings	$4,194,650.98

Table 3: Financial data statistics.

3.1.3 Identity Data

***Identity Data*: Measurement Methodology.** To assess the prevalence of potential identity data, we measure potential SSNs seen over the logged period. We check that the numbers fall within the issued range as listed by the Social Security Administration, but are unable to verify that the cards were already issued. Previous work has shown that an SSN is sufficient to steal an individual's identity, hence a publicly released SSN could put an individual at risk for identity theft [8].

Card Type	Counts
New	3,902
New In-Range	3,808
New Out-of-Range	94
Repeats	15,619

Table 4: Identity (SSN) statistics.

***Identity Data*: Measurement Results.** The results of our measurements are presented in Table 4. A total of 19,521 SSNs are identified or 0.15% of the corpus. In Section 3.1 we use random sampling to estimate that 0.40% of the messages in the corpus contain SSNs; again, this value is a reasonable estimate. The majority of potential SSNs are repeats with 3,902 unique values and 3,808 of these within the range of currently issued SSNs. We randomly sample around 3% of the unique in-range cards and cross-check them against the StolenIDSearch database. We found a single match. After inspecting the random sample, we found that 95% of the lines are explicitly labeled as SSNs. This finding suggests that the miscreants posting the cards believe the validity of the cards they posted, or are attempting to pass them off as valid.

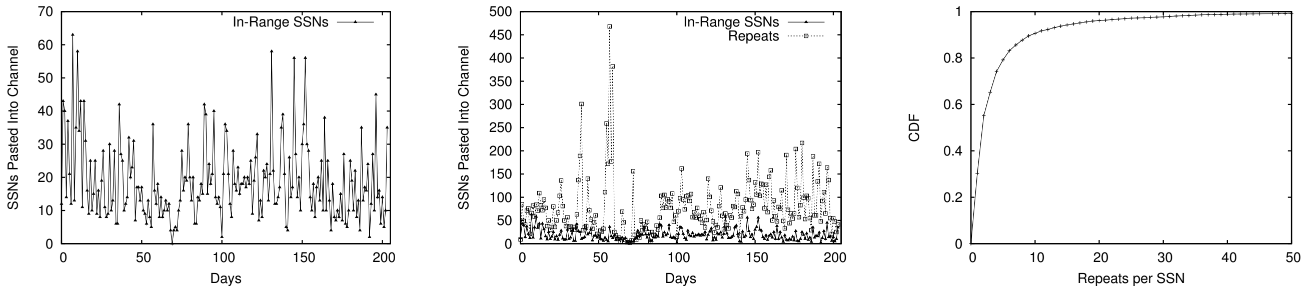
***Identity Data Rate*: Measurement Methodology.** In addition to establishing the number of SSNs in the channel and validating our prevalence estimate from Section 3.1, we measure the rate at which new in-range SSNs enter the channel and the rate at which previously seen cards are repeated.

***Identity Data Rate*: Measurement Results.** New in-range SSNs arrive at an average rate of 18.6 cards per day. Repeated SSNs arrive at an average rate of 76 cards per day. The majority of SSNs are repeated fewer than 3 times and 95% of cards are repeated 17 times or less. The results of our measurements are presented in Figure 5.

3.1.4 Estimating the Wealth of Miscreants

***Wealth*: Measurement Methodology.** To approach an estimate for the wealth stolen by the miscreants in this market, we add the potential losses from credit card fraud and financial account theft. Since the number of cards held in reserve is difficult to estimate, we use the number of cards with valid Luhn digits pasted to the channel. As an estimate for the amount of funds lost per card, we use the median loss amount for credit/debit fraud of $427.50 per card as reported in the 2006 Internet Crime Complaint Center's Internet Crime Report [6]. Our estimate assumes that all the card numbers with valid Luhn digits were active when posted to the channel and that the they incur an average loss of $427.50. We also assume that the financial accounts seen are valid and that all funds in the financial accounts are lost.

***Wealth*: Measurement Results.** With these estimates and assumptions, the total wealth generated from credit card fraud in the channel is over $37,000,000. If we include the financial account data from Section 3.1.2, we arrive at a total of over $93,000,000. While these numbers likely overestimate the wealth generated by the sensitive data posted to the channel, it is possible that market participants have many additional cards and financial accounts which they do not give away. This fact could make the estimated wealth established in this section only a fraction of the true value.



(a) New In-Range SSNs (b) New In-Range and Repeated SSNs (c) CDF of Repeats Per SSN

Figure 5: SSN arrivals and repetitions.

3.2 Market Participation

Having established the market being studied as an active market with significant levels of illegal activity, we shift our focus to the market's participants. We start by establishing a baseline activity level of the number of new and repeated messages posted per day. We measure the number of participants per day including new and old participants and the active lifetime of a participant over the logged period. We finish by correlating participant's IP addresses with IPs known to send spam, be infected with malware, or be open proxies.

3.2.1 Activity Levels

***Messages*: Measurement Methodology.** We measure the new and repeated messages per day. We manually checked outliers by randomly sampling to verify that the messages are the result of normal activity rather than message floods or other disruptive activity.

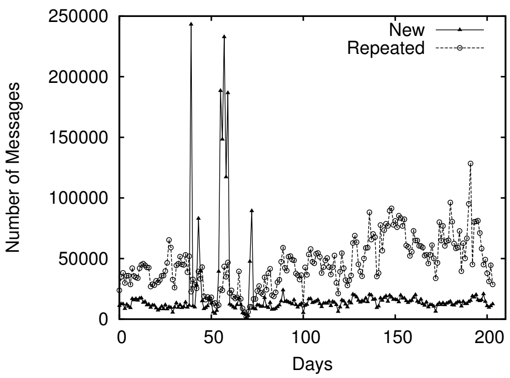


Figure 6: Message statistics.

***Messages*: Measurement Results.** Figure 6 shows the number of new and repeated messages per day. On average, over 64,000 messages are seen each day. The average number of new messages per day is greater than 19,000. After removing outliers corresponding to bursts of activity around day 50, the average rate of new messages per day drops to around 13,000. Repeated messages originate primarily from automated advertising scripts and arrive at an average rate of over 45,000 messages per day. These scripts repeat the same message at regular intervals to advertise the goods and services of sellers who may not be present at their terminals. Automated sales ads are common and on most days they comprise a majority of total channel messages.

3.2.2 Participant Identification

***Identifiers*: Measurement Methodology.** To assess the number of participants who contribute to the market each day, we measure the number of nicks (new and previously seen) who contributed at least one message to the market on a particular day. The number of nicks is not necessarily the same as the number of unique users since participants are not limited to using a single nick at a time. Scripts and automated bots also use nicks. In addition to the nicks seen in the logs, each market channel typically has a large number of lurkers who remain idle, sending zero public messages. These lurkers may be buyers who monitor channel ads and only contact sellers through private messages, leechers looking for free financial data, or fraud prevention services such as CardCops[2]. Our measurements do not include lurkers.

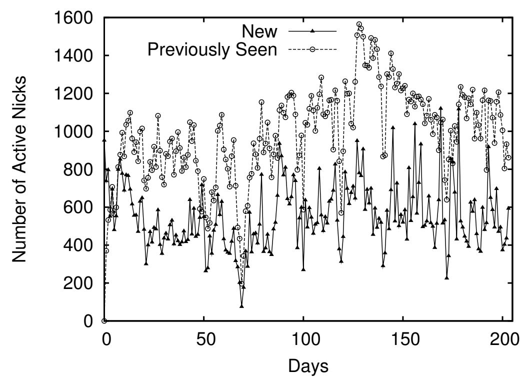


Figure 7: Participation by IRC nicks.

***Identifiers*: Measurement Results.** Figure 7 shows the results of our measurements. There were a total of 113,000 unique nicks seen over the monitored period. On an average day there are over 1,500 active nicks participating in the market. The majority of these nicks have been previously active in the channel at some time in the past. New nicks arrive at an average rate of 553 nicks per day.

***Active Lifetime*: Measurement Methodology.** Given the large number of previously seen nicks that operate in the channel, it is interesting to ask how long these nicks remain active. We measure the active lifetime of each nick, defined as the time between the nick's first and last message. Active lifetimes are useful to assess the extent to which participants build relationships by maintaining a nick over a long period versus creating new identities.

[2]http://www.cardcops.com

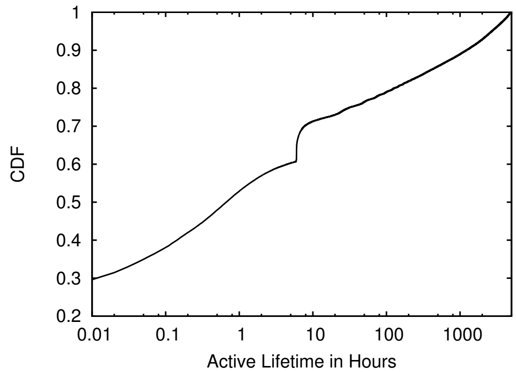


Figure 8: Active lifetime of IRC nicks.

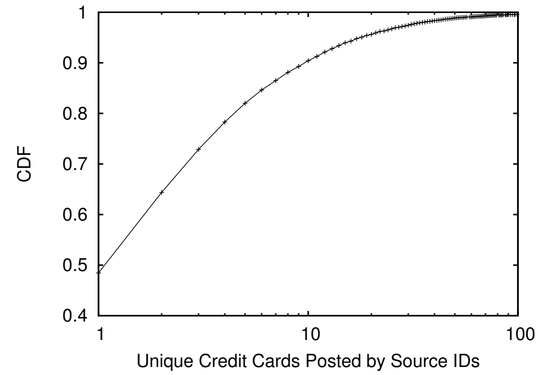


Figure 9: CDF of data samples posted by participants.

***Active Lifetime*: Measurement Results.** Figure 8 shows the active lifetimes of nicks on a logarithmic scale. 25% of all active nicks posted a single message to the channel, giving them an active lifetime of zero. The majority of nicks have a short active lifetime of less than 40 minutes while 95% of nicks have an active lifetime of less than 2,700 hours (112.5 days). The relatively long active lifetime of some nicks suggests that building relationships by maintaining a nick is a common and potentially lucrative practice.

***Client IP Lookups*: Measurement Methodology.** The second form of client identity that we measure is a client's IP address. We extracted a total of 65,513 IP addresses from the corpus and check the addresses against several blacklists. The first blacklist, the Spamhaus Block List (SBL)[3], is a "realtime database of IP addresses of verified spam sources and spam operations (including spammers, spam gangs and spam support services)." The second blacklist, the Exploits Block List (XBL), is a "realtime database of IP addresses of illegal 3rd party exploits, including open proxies (HTTP, socks, AnalogX, wingate, etc), worms/viruses with built-in spam engines, and other types of trojan-horse exploits." The XBL is composed of two lists: the Composite Block List (CBL)[4] and the NJABL[5] open proxy IPs list. Open proxies are commonly used to hide a client's IP address from law enforcement which may be monitoring a channel. The CBL catalogs IPs active in spam-related activities as a result of infection by bots or other malware.

BlackList	IPs On List	Percentage
XBL (CBL)	6,528	10%
XBL (NJABL)	939	1%
SBL	788	1%
–	60,305	90%

Table 5: Statistics from blacklist lookups.

***Client IP Lookups*: Measurement Results.** Table 5 shows the results of querying the blacklists. While 90% of the IPs are not on any blacklist, 10% are listed on the CBL suggesting that compromised hosts are being used to connect to the market. 1% of client IPs are on the SBL suggesting possible spamming activities and 1% are listed as open proxies.

[3] http://www.spamhaus.org
[4] http://cbl.abuseat.org/
[5] http://www.njabl.org/

3.2.3 Verified Status

The participants of the market operate in an environment of dishonesty and mutual distrust. Buyers and sellers must protect themselves from dishonest participants (a.k.a., "rippers") who purposely fail to uphold their end of a transaction. Such ripping behavior is common in other online markets and has lead to the establishment of reputation systems such as those found on eBay or Amazon Marketplace. Not surprisingly, establishing reputation in this underground market differs from traditional reputation establishment in online market places.

The primary mechanism to build credibility is by providing high-quality data "samples" which can be verified by a third party. The prevalence of free samples is part of what makes the existence of credit cards common in channel logs. After providing a sufficient number of verifiable samples of sensitive data, the channel's administrators consider a seller to be verified and give their nick a special designation, $+v$ (the 'voice' attribute), as a seal of approval. The validity of samples can be verified by performing a minimal cost transaction with the card, such as donating $1 to a charity of the miscreant's choice. The channel administration actively campaign for transactions to take place between verified participants – both sales and want ads carry notices that only transactions with verified participants will be accepted.

***Verified Status*: Measurement Methodology.** To better understand how a participant receives a verified status, we measure the number of credit cards posted by clients who provide at least one card during the monitored period. We use the client portion of the source identifier, including the nick and username, to distinguish clients. Results using other portions of the source identifier to distinguish client gave similar results.

***Verified Status* Measurement Results.** Figure 9 presents our results. The majority of clients who post sensitive data do so in small amounts and 95% post fewer than 18 samples. These measurements demonstrate suggest that participants, in particular sellers, need only post a small numbers of sensitive data samples to achieve verified status.

3.3 Market Services and Treachery

The channel service bot is an interactive script run by channel administrators for the purpose of providing useful services such as credit card limit checks and access to a BIN list. Table 6 describes commonly issued commands.

***Command Distribution*: Measurement Methodology.** We use syntactic matches to measure the number of times common commands were issued.

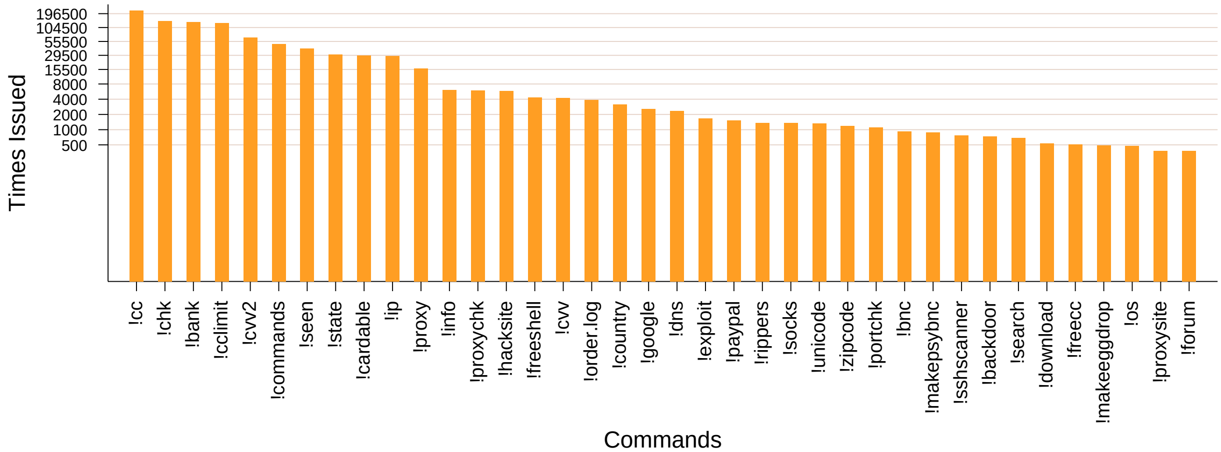


Figure 10: Command usage distribution.

***Command Distribution*: Measurement Results.** Figure 10 shows the distribution of command usage over the dataset. The top four commands are all associated with credit card data, either requesting data or purporting to provide card information.

Popular Commands	Meaning
!cc	Request for free credit card number
!chk $< CC >$	Request for valid or invalid status of $< CC >$
!bank $< BIN >$	Request for issuing back of cc with prefix $< BIN >$
!cclimit $< CC >$	Request for credit limit for $< CC >$
!cvv2 $< CC >$	Request for CVV2 of $< CC >$
!commands	Request for list of available commands
!seen $< nick >$	Request time $< nick >$ was last logged in
!state $< abbrev >$	Request full name for state $< abbrev >$
!cardable	Request for web merchant without card authorization check
!ip $< nick >$	Request IP address of nick $< nick >$
!proxy	Request for open proxy
!info	Request for general channel information
!proxychk $< addr >$	Request for status of proxy $< addr >$
!hacksite	Request for URL of hacking website

Table 6: Description of channel service bot commands.

A natural question to ask is what makes the risks associated with running this market worthwhile, or, equivalently, "What are the incentives for the market's administration?" While operating the market incurs a level of risk, it also provides an opportunity to easily acquire wealth. To understand the ease with which administrators may acquire wealth, one need look no further than the channel service bot commands. The channel services bot provides a number of commands which return information related to credit card numbers. Miscreants make constant use of these commands in an attempt to assess the wealth of their stolen data.

***Treachery*: Measurement Methodology.** After analyzing the source code for one channel services bot and looking through requests and responses in the corpus, we believe that the !chk, !cclimit, and !cvv2 are fallacious. For example, the !cclimit command parses the credit card number provided and returns a deterministic response without querying a database or attempting a transaction to infer the card's limit. One possible explanation for this finding is that the channel administrators run the channel bots as a way to steal credit card numbers from other participants. We measure the usage of one fallacious command to estimate the extent to which naive participants give away sensitive data. We check the card numbers provided as arguments to the command by performing a Luhn check and remove duplicates.

***Treachery*: Measurement Results.** Figure 11 shows the number of !cclimit commands issued which contain previously unseen card number. The command was issued a total of 129,464 times. We parsed responses to the command and found 25,696 unique cards, approximately one quarter of the total number of unique cards found in the corpus. These include 17,065 Visa, 6,705 Mastercard, 1,318 American Express, and 608 Discover cards.

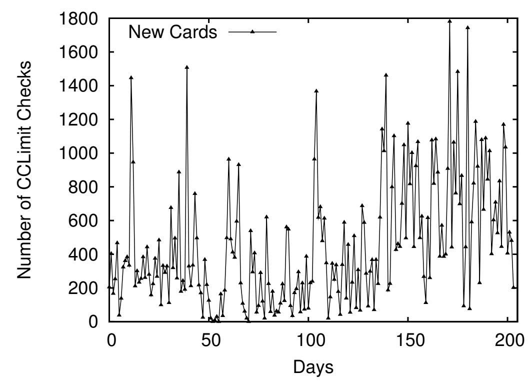


Figure 11: CClimit checks over time.

An average of 451 new cards are submit to the command per day. One expects the number of requests to the !cclimit command to decline over time as miscreants discover that the command provides a constant response over time. However, our measurements suggest that usage of the command is generally increasing over time. Possible explanations for this increase include that the data being submit to the !cclimit command is fake or new participants continuously join the channel and are tricked into using the command.

4. GOODS, SERVICES, AND PRICES

In this section, we measure the the number of sales and want ads for goods and services offered in the channel. The measurements in this section use both manual measurements and semantic measurements employing supervised machine-learning techniques.

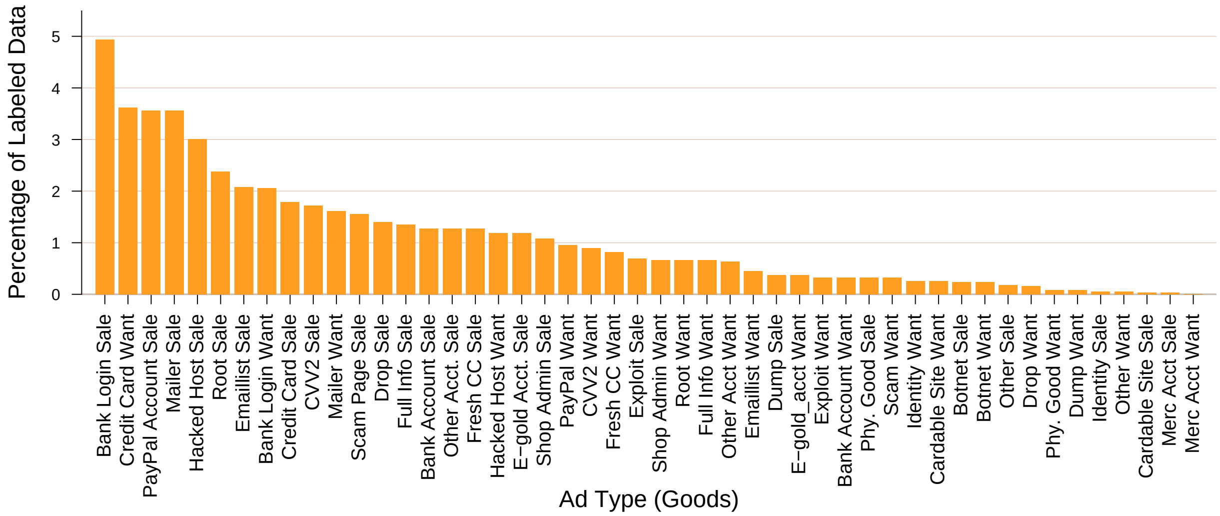


Figure 12: Distribution of ads for goods in labeled data.

4.1 Goods

***Ads for Goods*: Measurement Methodology.** Amongst the most common goods sold in the market are online credentials such as bank logins and PayPal accounts, sensitive data such as credit cards and SSNs, compromised machines, spamming tools including mailing lists and open mail relays, and scam webpages used for phishing.

***Ads for Goods*: Measurement Results.** Figure 12 shows the distribution of ads for goods from the labeled dataset. Sales ads outnumber want ads more than 2 to 1.

Having already established the extent to which sensitive data such as credit cards and SSNs constitute a large percentage of channel activity, we turn our attention to digital goods related to hacking, spamming and phishing. Ads for hacking-related goods include hacked hosts, root accounts, compromised e-merchant accounts, and software exploits. Ads for spam-related goods include web page email forms which can be used for spamming and bulk email lists. Ads for phishing-related goods primarily include scam webpages.

4.1.1 Hacking Related

***Hacking-Related Ads*: Measurement Methodology.** The most common hacking-related ads are those for compromised hosts. Sales ads for hacked hosts and root accounts constitute 5.39% of the labeled data while want ads for hacked hosts and root accounts constitute 1.85% of the labeled data. To determine the accuracy of these percentages as estimators for the percentage of compromised host want and sales ads for the entire corpus, we train two binary text classifiers to identify want and sales ads for compromised hosts. We train the classifiers using positive and negative examples of hacked host and root sales (want) ads from the training set, respectively.

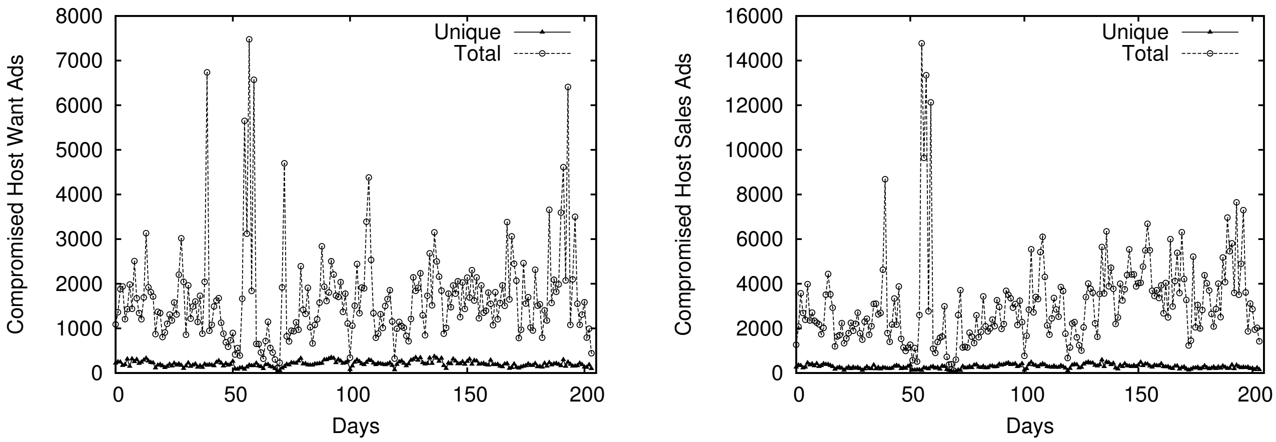
We evaluate the performance of both classifiers with the remaining 30% of labeled data in the test set. We report both precision and recall where $Precision = \frac{CorrectPositives}{PredictedPositives}$ and $Recall = \frac{CorrectPositives}{ActualPositives}$. We set the positive error penalization (-j option) to 3 and 8, respectively, to cause training errors on positive examples to outweigh errors on negative examples. This penalization was necessary to prevent the text classifier from achieving a high accuracy by always labeling messages as negative examples, erring only on the relatively infrequent positives examples. The compromised host sales ad classifier achieve a precision of 68.4% and a recall of 42.6%. The compromised host want ad classifier achieve a precision of 57.1% and a recall of 38.1%. In both cases, we chose classifiers with a higher precision and lower recall to limit the number of false positives. Higher recall percentages are possible if we allow for a lower precision, however this causes an inflation in the number of predicted positives. Even with their less than perfect classification accuracy, these classifiers efficiently filter the corpus and reduce the work required in subsequent analysis.

***Hacking-Related Ads*: Measurement Results.** We use the resulting text classifiers to label the 13 million unlabeled messages as either want ads for compromised hosts, sales ads for compromised hosts, or neither. We scale the resulting output by the $precision/recall$ ratio to roughly estimate the true positives in the corpus. When estimating values, we assume that errors are uniformly distributed over the dataset and that the error rates on the test set carry over to the entire corpus. Figure 13(a) shows the results of the want ad classifier and Figure 13(b) shows the results of the sales ad classifier.

The sales ad classifier identified an extrapolated 4.8% of the total corpus as sales ads for compromised hosts, with an absolute error of 0.59% from the previous estimate. The want ad classifier identified an extrapolated 2.6% of the total corpus as want ads for compromised hosts, with an absolute error of 0.75% from the previous estimate.

4.1.2 Spam and Phishing Related

As seen in Figure 12, the majority of spam and phishing-related ads in the labeled dataset are sales ads offering bulk email lists and sales offers for URLs of web email forms vulnerable to “email injection attacks.” An email injection attack exploits the input validation of web email forms such as the ubiquitous *contact us* form to include additional recipient email addresses. Rather than simply being sent to the individual responsible for the contact form, the web server sends the message to a list of injected addresses. The ease with which vulnerable email forms can be found has produced a bustling trade of such mailers. Mailer sales ads are the fourth most common type of ad for all goods, with bulk email list sales



(a) Want Ads (b) Sales Ads

Figure 13: Extrapolated number of ads for compromised hosts.

ads as the seventh most common. Vulnerable mailers ease the job of spammers who might otherwise have to locate open mail relays or employ bots to send spam. Email lists created by crawling webpages with email spiders or extracted from customer databases of compromised e-merchants further ease the job of spammers by enabling targeted spam campaigns.

4.1.3 Online Credentials and Sensitive Data

An extensive number of ads for online credentials from bank account logins to PayPal accounts were identified in the labeled data (See Figure 12.). In addition, want and sale ads for credit cards with associated information (cvv2, name, address, and answers to challenge questions) were common. Value-added features associated with credit card data include the freshness of the data and completeness of the associated information. Credit cards with cvv2 validation codes and full owner information which were recently acquired (fresh) garner a premium. Such cards are more flexible than cards with limited owner information or cards without their associated validation codes.

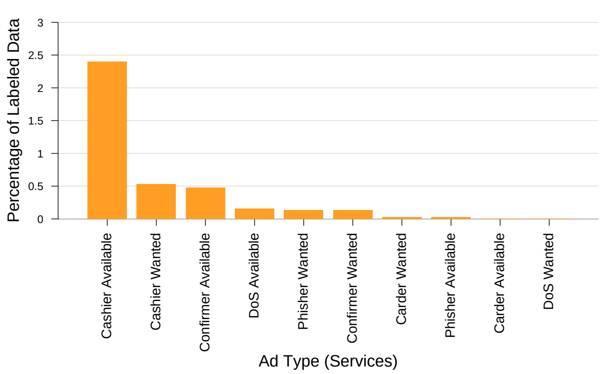


Figure 14: Distribution of ads for services in labeled data.

4.2 Services

***Ads for Services*: Measurement Methodology.** In addition to enabling access to various digital goods, the market includes a rudimentary offering of services primarily tailored to miscreants performing financial fraud. The most common service ad are offers for the services of a cashier, a miscreant who converts financial accounts to cash. Confirmers are used to assist in the verification step of Western Union money transfers. Money transfers from credit cards require a confirmation step where the individual transferring funds from the credit card answers questions to prove they are the card's rightful owner. This service is commonly offered on a gender-specific basis. In addition to financial fraud, a small percentage of service ads offer services such as DoS attacks, sending phishing emails, and purchasing goods with other's credit cards (a.k.a., *carding*).

***Ads for Services*: Measurement Result.** Figure 14 shows the distribution of service ads over the labeled dataset.

4.3 Prices

Before public underground markets were established, quantifying the cost or difficulty of obtaining a compromised host, a spam relay, or an identity was highly subjective. One might estimate the cost by performing calculations which depend on opaque quantities such as an attacker's prowess or level of qualitative skill level such as "script kiddie" or "elite hacker." Such qualitative techniques rarely meet the requirements of organizations seeking to assess their exposure to security-related risks or researchers interested in measuring the security of a system. Given that active underground markets exist with individuals buying and selling goods and services of all types, we can monitor these markets to quantify the difficulty of acquiring a resource. In particular, underground markets establish the monetary cost to acquire an illegal good such as a compromised host.

To demonstrate why quantifying the difficulty of acquiring a resource in monetary terms is useful, consider the case of a DDoS defense scheme. A useful technique to evaluate such schemes is the number of machines (or the level of resources) required to overwhelm the defense. DDoS defense papers are rife with claims of security against various resource levels; however, they often fail to quantify the cost of acquiring such resources. The machine learning techniques and analysis in this paper can fill this void by establishing prices for relevant goods and services.

***Price of Compromised Hosts*: Measurement Methodology.** We first extract all messages which contain explicit prices (a dollar sign and at least one non-zero digit) and remove repeated messages. Next, we use the SVM classifier trained to identify sale ads for compromised hosts to filter the remaining lines for just those lines containing asking prices for compromised hosts. Finally, we randomly sample the asking prices and manually extract prices.

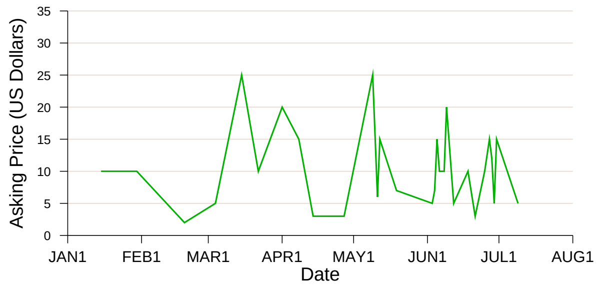


Figure 15: Price schedule for compromised hosts.

***Price of Compromised Hosts*: Measurement Results.** The results of our measurements are presented in the price schedule for compromised hosts in Figure 15. This schedule enables defenders to establish the *cost to buy* sufficient resources to overcome a defense system. For example, a DDoS defense that is effective up to 1,000 hosts can be overwhelmed by $100,000 in January or as little as $20,000 in February. The *cost to buy* can be used to assess the strength of an adversary with resources r at time t. For example, a $10,000 adversary can purchase 1,000 compromised hosts in January. For simplicity, we assume that that sufficient quantity is available to satisfy the quantity demanded.

In addition to measuring the cost to buy compromised hosts, the measurement techniques used in this paper can assist in measuring costs for resources used by spammers, phishers, and identity thieves. These prices can be used to establish the cost to send targeted spam emails, to purchase a bank or PayPal account, or to steal an identity. Further evaluation is necessary to validate that the cost to buy a resource provides an accurate and reliable metric to measure the risk associated with a resource when an adversary's resources are expressed in monetary terms.

5. DISCUSSION

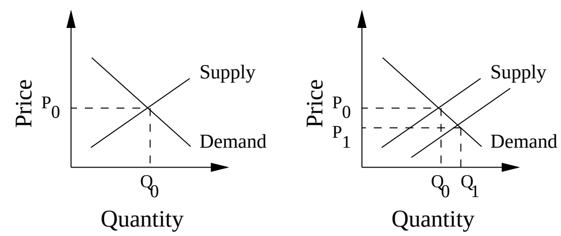
We begin with a discussion of how the market data gathered in this paper could potentially provide a new approach for quantifying Internet security. Next, we discuss potential low-cost approaches to disrupt the underground markets which deviate significantly from one approach currently used by law enforcement [17]. Although the approaches we describe in this section rely on oversimplifications, we believe that our preliminary explorations will help motivate the challenges that lie ahead and encourage further research.

5.1 Inferring Global Statistics and Trends

Measuring global statistics and trends such as the number of compromised Internet hosts or the number of stolen identities is a difficult task. Not only do these phenomena exhibit a significant variance over time, but they are difficult to directly measure due to insufficient coverage.

We consider the task of measuring trends in the total number of compromised hosts on the Internet. We take an economic approach to measurement which deviates significantly from previous, primarily statistical approaches. Rather than measuring the number of packets received at a network telescope and extrapolating the aggregate number of compromised hosts based on a random-scanning assumption, we can use the laws of supply and demand and market measurements to infer global trends.

The law of supply states that, all other factors remaining constant, the supply of a good or service is proportionate to its price. The law of demand states that, all other factors remaining constant, the demand for a good or service is inversely proportionate to its price. These laws establish the well-known supply and demand curves shown in Figure 16(a). When we observe the equilibrium price for a good or service in a market, the price provides the y-coordinate of the intersection point of the supply and demand curves. As the supply and demand curves shift in response to market forces, we observe changes in the market equilibrium price. Given that we are unable to directly measure the quantity of goods or services available, we need a method to infer quantities or the change in quantity supplied or demanded at a point in time. If we assume that demand remains constant over short time periods – or we establish that demand has remained constant by directly measuring the forces which cause shifts in demand (population, income, price of a substitute or complement, and expected future value) – then changes in the price of a good or service are the result of supply-side factors. An example of an increase in supply and the corresponding effects is illustrated in Figure 16(b).



(a) Supply and demand curves

(b) Shift of supply curve.

Figure 16: Inferring underlying market trends.

Consider the price of a compromised host in an underground market. Under the assumption of constant demand, if the equilibrium price for a compromised host at time t is P_0 and the price is P_1 at time t+1, then we can infer that the total quantity of compromised hosts available has increased. Even if we aren't able to directly measure the quantities Q_0 or Q_1, the laws of supply and demand provide us with the ability to measure trends. A similar analysis is possible when supply remains constant and demand-side factors cause a shift in the demand curve. In addition, more sophisticated econometric techniques such as simultaneous equation models can be used to solve for supply and demand curves.

5.2 Efficient Countermeasures

Underground markets represent a substantial security threat. Previous approaches for disrupting underground markets have focused on standard law enforcement activities such as locating and disabling hosting infrastructure or identifying and arresting market participants [17]. These techniques face numerous social and technological hurdles which limit their success and result in substantial associated costs. For example, disabling the hosting infrastructure for a market may require multi-national cooperation, which can be time and resource consuming. Furthermore, nations may refuse to cooperate with foreign law enforcement agencies or may lack appropriate laws for prosecution. Even in the case where law enforcement techniques have succeeded in disrupting an underground market, the markets often re-emerge under new administration with a new "bulletproof" hosting infrastructure. Identifying and arresting key players also includes a host of associated complexities and costs, such as tracing individuals through chains of compromised hosts and the cost of subsequent legal proceedings.

The substantial costs and limited success of standard law enforcement techniques motivate our search for low-cost approaches to countering the threat posed by underground markets.

In this section, we sketch two low-cost countermeasures based on principles in economics and natural limitations in the client identification capabilities of open underground markets. The first is a Sybil attack and the second is a slander attack. The goal of this preliminary exploration is to highlight open challenges and present initial approaches on how to tackle them.

5.2.1 *Sybil Attack*

In a Sybil attack on a voting system, an attacker creates numerous identities (Sybils) in order to control a disproportionally high percentage of votes [7]. Using a similar idea, we can exploit the open nature of the underground market to establish Sybil identities which in turn disrupt the market by undercutting its participant verification system. To demonstrate our attack concretely, we describe it in the context of the market studied in this paper. Our attack proceeds in three stages: 1) Sybil generation, 2) achieving verified status, and finally 3) deceptive sales.

Sybil Generation. In stage one, an attacker establishes multiple Sybil identities by connecting to the market's IRC servers and registering nicknames. The required number of Sybil identities depends on the number of verified-status sellers in the market. A higher ratio of Sybils to verified-status sellers will improve the overall effectiveness of the attack.

Achieving Verified Status. In stage two, an attacker builds the status of each Sybil identity. This can be accomplished through positive feedback from other Sybils or out-of-band activities. The verification system of the underground market studied in this paper provides verified status to participants who freely provide high-quality credit card data. The success of a Sybil attack depends on the cost associated with generating a Sybil identity and the cost of achieving verified status. For a Sybil attack to be successful, these costs must be minimized. For the studied market, a low-cost technique to achieve verified status is to enter several separate IRC channels and replay credit card data seen in one channel to a different channel. This allows verified-status Sybils to be produced at a minimal cost.

Deceptive Sales. In stage three, an attacker utilizes their verified-status Sybils to advertise goods and services for sale. Rather than undergoing an honest transaction, the attacker first requests payment and subsequently fails to provide the good or service promised. Such behavior is known as "ripping" and it is the goal of the verification system to minimize such behavior. However, poor controls on how one achieves verified status and establishes identities make it possible to undermine the market's verification system. If an attacker's Sybils are indistinguishable from other verified-status sellers, a buyer will be unable to identify honest verified-status sellers from dishonest verified-status Sybils. In the long term, buyers will become unwilling to pay the high asking price requested by verified-status sellers because of the buyer's inability to assess the true quality of sellers.

Markets which exhibit this form of asymmetric information, where buyers are unable to distinguish the quality of goods, are known as *lemon markets* [2]. Lemon markets see a reduction in successful transactions until the information asymmetry is corrected. In our case, the market would need to establish a verification system which is robust against Sybil attacks. One approach would be to detect anomalous recommendation topologies [21], but this would require a sophisticated system for tracking interactions over time. Another approach would be to increase the cost of establishing an identity, in turn pushing the market towards a closed market, which discourages new individuals from joining – subsequently raising the barrier to entry for cybercrime.

5.2.2 *Slander Attack*

In a slander attack, an attacker eliminates the verified status of buyers and sellers through false defamation. By eliminating the status of honest individuals, an attacker again establishes a lemon market. To understand why, consider a market with one verified-status seller, Honest Harry, one unverified seller, Dishonest Dale, and an unlimited number of buyers. If the verification system accurately classifies individuals into honest and dishonest classes, in turn minimizing the variance in expected payoff of an entity, Honest Harry will charge a premium for his goods since a buyer's expected payoff when undertaking a transaction with Harry will be higher than their expected payoff with Dishonest Dale. Assume a number of buyers slander Harry, subsequently eliminating his verified status. As a result, buyers will lower their expected payoff for transactions with Harry under the assumption that Harry is less honest than before (exhibits a higher variance in payoffs). However, having remained honest, Harry will be unwilling to accept a lower price (since in an efficient market Harry is already selling at equilibrium). Buyers will, in turn, leave the market or undertake transactions with Dishonest Dale, who may fail to uphold his end of a transaction. Regardless, the result is a marked decrease in the number of successful transactions – a desired outcome.

6. RELATED WORK

Related work falls into two categories: underground markets and the economics of information security.

Previous studies have framed the existence of underground cyber markets, but have not systematically analyzed the markets [18, 19]. We employ machine learning techniques and random sampling to classify logs into a number of categories; allowing us to assess the extent of miscreant behavior rather than only observing snapshots in phenomenological terms.

Anderson discusses why security failures may be attributable to "perverse economic incentives" in which victims bear the costs of security failures rather than those who are responsible for the system's security in the first place [3]. Schechter develops an argument that the cost to break into a system is an effective metric to quantitatively assess the security of the system [15]. Schechter also suggests that vulnerability markets could be set up to entice hackers to find exploits. The lowest expected cost for anyone to discover and exploit a vulnerability is the *Cost to Break* metric. Schechter also advances an econometric model of the security risk from remote attack [16]. In comparison, this paper proposes a security metric not for a particular system with unknown vulnerabilities, but for the Internet as a whole. Similar to the cost to break metric, our proposed metric uses market pricing. Ozment reformulated Schechter's vulnerability markets as "bug auctions" and applied auction theory to tune market structure [14]. In a position paper, Aspnes et al. state as a key challenge that of obtaining quantitative answers to the scope of Internet insecurity [4]. Aspnes et al. also state that "economics provides a natural framework within which to define metrics for systemic security." The approach proposed in this paper hopes to partially fulfill this goal.

7. CONCLUSION AND FUTURE WORK

Internet miscreants of all sorts have banded together and established a bustling underground economy. This economy operates on public IRC servers and actively flaunts the laws of nations and the rights of individuals. To elucidate the threat posed by this market, we performed the first systematic study including extensive measurements of 7 months of data and the use of machine learning techniques to label messages with their associated meanings.

To stimulate further research, we discussed how our measurements might be applied to quantify the security of systems and to estimate global trends that are difficult to measure, such as changes in the total number of compromised hosts on the Internet. Further, we sketched efficient, low-cost countermeasures which use principles from economics to disrupt the market from within. These countermeasures deviate significantly from today's use of law enforcement or technical approaches, which meet with substantial costs.

The ready availability of market data for illegal activities begs a number of interesting questions. For example, how does the market respond to security-related incidents such as the discovery of an exploit or the release of a patch? The use of economic event studies may allow us to better understand the true costs and benefits of deployed security technologies, data breeches, and new security protocols. In addition to studying effects, tracking underground market indices may allow for accurate forecasting and predictions of the future state of Internet security. We consider this study an initial step towards the use of economic measurements of underground markets to provide new directions and insights into the state of information and Internet security.

8. ACKNOWLEDGEMENTS

This work would not have been possible without help from Team Cymru, Alessando Acquisti, Paul Bennett, John Bethencourt, Gaurav Kataria, Chris Kanich, Leonid Kontorovich, Pratyusa K. Manadhata, Alan Montgomery, Amar Phanishayee, Srini Seshan, Jeff Terrell, and Rob Thomas. We also wish to thank the anonymous reviewers and numerous anonymous miscreants.

9. REFERENCES

[1] Email injection. http://www.securephpwiki.com/index.php/Email_Injection, August 2007.

[2] George A. Akerlof. The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970.

[3] Ross Anderson. Why Information Security is Hard - An Economic Perspective. In *17th Annual Computer Security Applications Conference*, 2001.

[4] J. Aspnes, J. Feigenbaum, M. Mitzenmacher, and D. Parkes. Towards better definitions and measures of internet security. In *Workshop on Large-Scale Network Security and Deployment Obstacles*, 2003.

[5] Paul N. Bennett and Jaime Carbonell. Feature Representation for Effective Action-Item Detection. In *ACM SIGIR Special Interest Group on Information Retrival*, 2005.

[6] Internet Crime Complaint Center. Internet crime report. http://www.ic3.gov/media/annualreport/2006_IC3Report.pdf, Jan. - Dec. 2006.

[7] John R. Douceur. The sybil attack. In *Proceedings of the IPTPS Workshop*, 2002.

[8] Serge Egelman and Lorrie Faith Cranor. The Real ID Act: Fixing Identity Documents with Duct Tape. *I/S: A Journal of Law and Policy for the Information Society*, 2(1):149–183, 2006.

[9] Thorsten Joachims. *Advances in Kernel Methods - Support Vector Learning*, chapter Making Large-Scale SVM Learning Practical. MIT-Press, 1999.

[10] Michael C. St. Johns. Identification protocol. RFC 1413, February 1993. http://tools.ietf.org/html/rfc1413.

[11] Hans P. Luhn. Computer for verifying numbers. U.S. Patent 2,950,048, August 1960.

[12] Charlie Miller. The legitimate vulnerability market: Inside the secretive world of 0-day exploit sales. In *Sixth Workshop on the Economics of Information Security*, May 2007.

[13] Jarkko Oikarinen and Darren Reed. Internet relay chat protocol. RFC 1459, March 1993.

[14] Andy Ozment. Bug Auctions: Vulnerability Markets Reconsidered. In *Third Workshop on Economics and Information Security*, 2004.

[15] Stuart E. Schechter. Quantitatively Differentiating System Security. In *First Workshop on Economics and Information Security*, 2002.

[16] Stuart E. Schechter. Toward econometric models of the security risk from remote attack. *IEEE Security and Privacy*, 03(1):40–44, 2005.

[17] United States Secret Service. United states secret service's operation rolling stone nets multiple arrests: Ongoing undercover operation targets cyber fraudsters. Press Release, March 2006.

[18] DeepSight Analyst Team. Online Fraud Communities and Tools. Technical report, Symantec, January 2006.

[19] Rob Thomas and Jerry Martin. the underground economy: priceless. *USENIX ;login:*, 31(6), December 2006.

[20] Yiming Yang and Xin Liu. A Re-examination of Text Categorization Methods. In *ACM SIGIR Special Interest Group on Information Retrival*, 1999.

[21] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: Defending against sybil attacks via social networks. In *Proceedings of ACM SIGCOMM*, August 2006.