Scambaiter: Understanding Targeted Nigerian Scams on Craigslist

Youngsam Park* Jackie Jones† Damon McCoy† Elaine Shi* Markus Jakobsson‡

*Department of Computer Science
University of Maryland, College Park
{yspark, elaine}@cs.umd.edu

†Department of Computer Science
George Mason University
jjones24@masonlive.gmu.edu, mccoy@cs.gmu.edu

‡ZapFraud

***Abstract*—Advance fee fraud scams, also known as Nigerian scams have evolved from simple untargeted email messages to more sophisticated scams targeted at users of classifieds, dating and other websites. Even though such scams are observed frequently, the community's understanding of targeted Nigerian scam is limited since the scammers operate "underground". In this paper, we focus on fake payment scams targeting users on Craigslist. To better understand this type of scam and associated scammers, we built an automated data collection system. The system relied on what we term *magnetic honeypot advertisements*. These are advertisements that are designed to *attract scammers but repel legitimate users* – similar to how a magnet attracts one side of a magnet but repels the other. Using advertisements of this type, we offered goods for sale on Craigslist, gathered scam emails and interacted with scammers. We use this measurement platform to gather three months of data and perform an in-depth analysis. Our analysis provides us with a better understanding of scammers' action patterns, automation tools, scammers' email account usage and distribution of scammers' geolocation. From our analysis of this dataset, we find that around 10 groups of scammers were responsible for nearly half of the over 13,000 total scam attempts we received. These groups use shipping address and phone numbers in both Nigeria and the U.S. We also identify potential methods of deterring these targeted scams based on patterns in the scammer's messages and usage of email accounts that might enable improved filter of their initial messages by content and email address.**

I. Introduction

Advance fee fraud, more commonly referred to as *Nigerian scams* or *419 scams*[1], is a prevalent form of online fraud that not only causes financial loss to individuals and businesses alike [3], but also can bring emotional or psychological damage to victim users [19]. An estimation of global losses to Nigerian scams in 2005 is more than 3 billion dollars [14]. This scam was originally mostly untargeted and delivered via email spam. However, today there are more sophisticated targeted versions of this scam that are directed at users of classifieds, jobs and dating sites.

In spite of its prevalence, the community's understanding of targeted online Nigerian scams is still lacking. Many online websites, such as Craigslist, filter out scam postings to protect its legitimate users. For example, Craigslist has many safeguards in place to prevent scam postings, such as requiring phone number verification for a Craigslist account to prevent scammers from registering large numbers of Craigslist accounts and posting fraudulent advertisements, blocking suspicious IP addresses and accounts, and removing advertisements containing suspicious content. However, little is done to protect users from receiving scam replies to their advertisements. In addition, email service providers face a significantly more challenging task when attempting to filter lower volume and target advance fee fraud spam rather than less targeted and more common spam (e.g., pharmacy campaigns).

In this paper, we focus on Nigerian scams on Craigslist, one of the most popular online market websites whose monthly visitors are over 60 million in the U.S. alone[2]. We present an in-depth measurement study of such scam activities. Through this measurement study, we aim to better *understand the underground economy of Nigerian scams*, and *seek effective intervention points*. In particular, we seek to address questions such as the following: "Where are scammers located?", "How do scam factories operate?", "What features can we use to distinguish a scam email from a legitimate email?"

In order to better understand Nigerian scams on Craigslist, we posted magnetic honeypot advertisements – designed to attract scammers but repel legitimate users. We received and replied to scam emails resulting from our advertisements, and analyzed the emails. For quantitative analysis of scams, we build an automated data collection system which posts advertisements, collects scam emails and interacts with scammers by sending out a response to the received scam emails. We also collect IP addresses of scammers to explicitly confirm geolocation of the scammers. We perform various analysis of the massively collected dataset to better understand how scammers work. We also cluster observed scammers into groups based on a few key factors such as email addresses, shipping address, phone number and email payload.

Our analysis reveals that these types of Nigerian scams are highly prevalent as our magnetic honeypot advertisements

[1] We use all three terms interchangeably in this paper.

[2] http://www.craigslist.org/about/factsheet

NDSS '14, 23-26 February 2014, San Diego, CA, USA
Copyright 2014 Internet Society, ISBN 1-891562-35-5
http://dx.doi.org/doi-info-to-be-provided-later

on average received 9.6 scam replies. The most enlightening result of our analysis is that *about 50% of the scam attempts observed can be linked back to the top 10 groups*. These groups are targeting advertisements spread over many classes of goods and geographic regions of Craigslist. In addition, our analysis reveals that many of the initial scam messages are automated and arrive from a large number of email address that are quickly abandoned. However, most of these initial messages contain a different reply-to address to a smaller set of longer lived email accounts. We also find that 23% of the shipping addresses are located in the United States, although most of the IP addresses and shipping addresses are located in Nigeria. This indicates there are likely either accomplices or reshipping operations being used. Our analysis of the content of the messages shows certain occurrences of words such as, God, overseas military personnel, and capital letters that might be used to help filter these messages.

From this analysis we find several potential intervention points. Our analysis of the message content indicates that message filtering could be improved by looking for combinations of these pattern such as a reply-to address that does not match the sender's address, usage of these uncommon phrases, and identification and blacklisting of these more stable and long-lived secondary accounts. Also, shipping addresses might be the starting point for law enforcement investigations. Along these same lines the fact that only ten groups of scammers accounted for nearly half of the scams we received indicates that it might be possible to target and disrupt these groups, greatly reducing the prevalence of this scam.

II. Related Works

There have been a number of previous studies that have looked at the structure by Smith [14], Buchaman and Grant [1] and estimated losses from advance fee fraud by Dyrud [2]. Whitty and Buchaman [19] and Rege [13] have investigated the dynamics of online dating scams. More broadly, Stajano and Wilson [15] created a taxonomy of the different types of psychology motivations used by scammers. Garg and Nilizadeh [5] investigated whether economic, structural and cultural characteristics of a community affects the scams on Craigslist. Their work focuses on potential scammers' advertisements posted on Craigslist. Tive [18] introduced in his study various techniques of advance fee fraud. Herley [7] has argued that Nigerian scammers deliberately craft their messages to be unbelievable as a method of reducing the number of replies from people that are unlikely to fall victim to these scams. In contrast, our study aimed to be more focused on collecting empirical data to enable a data-driven analysis that does not rely in self reported statistics. Isacenkova et al. [8] identified a thousand scam groups from an existing scam email dataset with the help of a multi-dimensional clustering technique. This study also argued that scammers' email addresses and phone numbers are crucial factors of the clustering. Goa et al. [4] investigates the use of ontology-based knowledge engineering for Nigerian scam email text mining. Unlike previous studies, in our investigation we have focused on 1) understanding in great depth the prevalence and techniques, and 2) identifying the structure of larger scale groups of scammers that are engaged in attempting to defraud people posting goods for sale on Craigslist.

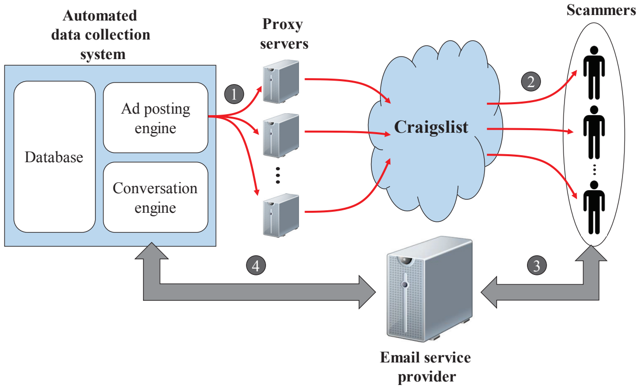


Fig. 1. **Automated collection scam data using magnetic honeypot ads.**
(1, 2): The system posts "magnetic honeypot" ads which would attract scammers only;
(3): the scammers send scam emails in response to the magnetic honeypot ads;
(4): the system automatically engages in email conversations with scammers.
Fraud attempt: The conversation eventually leads to a fraud attempt, where the scammer sends a fake PayPal notification or fake check, and urges the victim to send the goods to the scammer-indicated mailing address.

Another large body of recent work has set about conducting empirical measurements to understand the dynamics and economic underpinnings of different types of cybercrime. Much of this work has been focused on spam email [9], [16], illicit online pharmacies [11], and mapping out scam hosting infrastructure [10], [17]. Our work builds on this, but focuses deeply on the Nigerian scam problem in particular. We have conducted, to our knowledge, the first large scale empirical measurement study of 419 scams. It provides us with insights into how these scams are organized and how they might be better deterred in the future.

III. Data Collection Methodology

We have built an automated data collection system that collects scam data on Craigslist as illustrated in Figure 1. Our data collection methodology is explained below.

A. *Creating magnetic honeypot posts*

Our data collection focused on selling a variety of goods on Craigslist.

Our idea is to create *magnetic honeypot* advertisements that would *selectively attract scammers but not legitimate users*. To do this, we post unattractive advertisements, e.g., selling a used iPad at a price higher than new. More specifically, we choose goods among a list of popular items on Amazon to make sure that the goods we are selling can be easily bought from Amazon or anywhere else. The selling price is set to be a little bit higher than the price of new product found on Amazon. Any sensible real user would conceivably not reply to such posts. However, scammers would — they might be using bots to crawl Craigslist or automate the response process, or might not carefully check the contents of each post due to lack of labor.

We made sure that our data collection methodology conforms to good ethical standards, as further discussed in Section III-D.

B. Automated communication with scammers

We have built an automated conversation engine that performs linguistic analysis of incoming emails from scammers, and automatically engages in multiple rounds of communication with scammers. The engine periodically checks inboxes of email accounts used for Craigslist accounts and reads in all unread emails. Then it classifies the emails to identify valid scam emails. Our automated engine replies to a subset of the scam emails we receive — specifically, emails with a subject line that replies directly to the subject of our post. Henceforth, our automated engine exchanges multiple rounds of emails with the scammer, leading to the fraud attempt, e.g., fake PayPal notifications or fake checks. The most common type of fraud we observe is a fake PayPal notification stating that funds have arrived at the victim's PayPal account, followed by requests for the victim to send the product to the scammer's mailing address.

A typical example of email conversation is posted in Figure 2, and more examples are posted in Appendix A.

iPhone 5 64GB (WashingtonDC)

[from: cathy caraballo <cathycaraballo93@gmail.com>]

```
Still available for sell??
```

[Our response]

```
Yes, the product is still available. Please let me
know if you need more information.
```

[from: cathy caraballo <cathycaraballo93@gmail.com>]

```
Thanks for getting back to me [words omitted] l
will give you $680 for the item in order to out
bid other buyer and $60 for shipping via a register
mail down to my Son,kindly get back to me with
your PayPal email account so l can proceed now with
your payment and if you don't have an account with
PayPal, its pretty easy, safe and secured to open
one. Just log on to WWW.PayPal.com [words omitted]
Thanks and God Bless.
```

[Our response]

```
Sounds great. My paypal account is
sarkadejan@gmail.com. Thanks!
```

[from: cathy caraballo <cathycaraballo93@gmail.com>]

```
Hello Friend.just want you to know that your
payment has been made paypal just mailed me now
so check your inbox or spam and your money has been
deducted from my account pending to your account..
[words omitted] tracking number and scanned receipt
for verify and Here is the Shipping Details below
Name..xxx xxxxx
address..xxx xxxxx st
city,Bakersfield
state..california
zipcode.93307
Best Regard ..
```

Fake Paypal notification:
[from: service@paypal
<verifedtrackingshipp@mail2consultant.com>]

```
Dear Sarkadejan@gmail.com, You've received
an instant payment of $770.00 USD from Cathy
Caraballo93, [words and images omitted]
```

Fig. 2. **Example 419 scam thread.** The first scam response usually has one or couple of simple sentences showing scammer's interest in goods posted by the victim. The second scam response contains a fraud attempt through fake PayPal or bogus check. The scammer's offer is usually attractive since their offer price is higher than then victim's list price. Finally, the third and later scam responses urge the victim to send the goods to the designated mailing address.

C. IP address collection

The IP address of an email sender provides insightful information, such as scammers' geolocation. However, collecting IP addresses from email headers is not always feasible when the emails are relayed by the site (e.g., Craigslist), or if the webmail provider does not include source IP address in email headers (e.g., Gmail). To collect IP addresses of scammers, our automated conversation engine embeds an external image link into emails generated in response to a received scam email. Since the embedded link leads to a web server under our control, we can collect IP addresses of anyone who accesses image files we've embedded. The embedded link is unique to the corresponding advertisement so that we can later analyze the collected IP addresses based on factors such as city, product category and price.

D. Ethics

Since our experiment ultimately deals with human subjects, we put several controls in place to manage any harm to the participants. In addition, we went through the process of getting our experiment approved by our institution's human subjects review process. During the experiment, we collected scam emails by posting honey pot advertisements which may attract responses from legitimate users as well as scammers. Even though our honey pot advertisements are designed to be "unattractive" such that legitimate users would not be interested in replying, it is still possible that our experiment might receive responses from legitimate users that send an actual payment to buy a product that we have posted on Craigslist. In order to prevent this unintentional "victimization", we consistently check if there were any actual payments made by legitimate Craigslist users. If a payment was made by a legitimate user, the victim would be provided with pertinent information about our experiment, and the item would be shipped to them or the refund procedure would be initiated immediately. In addition, any messages from this user would be purged from our collected data. Note that fortunately, we found no payment made by any legitimate users during the entire experiment.

Another issue concerns how we use the collected data that might contain private information about scammers. Throughout the experiment, we gathered messages that contain information such as shipping addresses and phone numbers which could potentially be used to identify scammers. We limit the use of raw data to email addresses, IPs, and text from messages that will not clearly identify the actual identity of the scammer. All other information is only included in aggregate to avoid revealing the identity of any scammers.

Finally, we adhered to Craigslist's terms of use regarding posting advertisements [3]. Specifically, each of our accounts only posted in a single location and were restricted to a posting rate of once every 48 hours.

[3] http://www.craigslist.org/about/terms.of.use

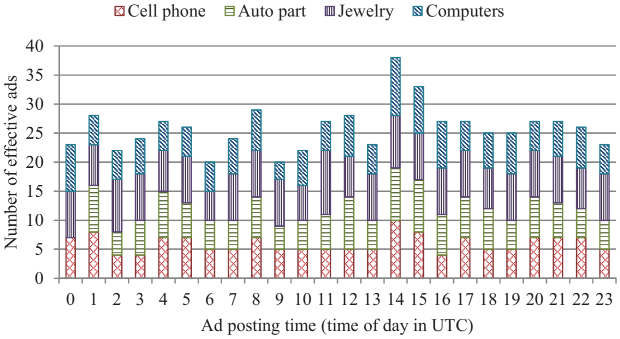


Fig. 3. **Distribution of magnetic honeypot ads over ad posting time.** The ad posting engine posts magnetic honeypot ads every 48 hours or more in each city and category.

IV. Experimental Results

In Sections IV and V to follow, we first present a summary of the dataset we collected and our findings of this measurement study.

A. Dataset

Table I presents a summary of the dataset we collected using the methodology described in Section III. More details of each part of the table are explained below.

1) Overview and terminology: Our 419 scam data collection spans a duration of roughly *three months*, from 4/15/2013 to 7/19/2013. We selected 20 locations including 10 large and 10 small cities/areas from a list provided by Craigslist. The large cities include San Francisco, Seattle, New York, Boston, LA, San Diego, Portland, Washington DC, Chicago and Denver and small cities/areas include Twin Tiers, Cumberland Valley, Meadville, Susanville, Siskiyou, Hanford-Corcoran, Santa Maria, Winchester, Southwest and Eastern Colorado.

We selected four product categories including *cell phone*, *computer*, *jewelry* and *auto parts*, which are used by many Craigslist users and therefore, many advertisements are posted daily posted as usual. As mentioned in Section III-D, we posted our ads at very low rates, so that they account for only an unnoticeable fraction of the total traffic volume in each city on Craigslist. Specifically, we posted at most one advertisement per category per city every 48 hours, which makes at most 80 advertisements per 48 hours in total. The price of products used in the experiments ranged from $80 to $7,000.

Table II shows the terminology that we use to refer to honeypot ads and received emails throughout the paper.

Effective ads	Magnetic honeypot ads that are not flagged by Craigslist until the expiration (1 week)
Email thread	Several emails in the same conversation
First response received	The first email sent by the scammer to us after seeing our Craigslist ads.
First reply sent	Our response to the first response received.
Second response received, second reply sent	The scammer's response to our first reply; our reply to that in turn

TABLE II. **Terminology**

2) Magnetic Honeypot Advertisements: During the experiment, we posted $1,376$ magnetic honeypot advertisements over 20 large and small cities in the U.S. Among the whole advertisements posted, 747 advertisements were flagged by Craigslist, leaving 629 *effective* advertisements. 42 emails accounts (Craigslist accounts) were used during the experiment. We designed our system to post magnetic honeypot advertisements evenly distributed over posting time and product category to minimize possible biases in the collected dataset. Figure 3 illustrates distribution of effective advertisements over time of day. In this figure, the slight unevenness in distribution (over different times of the day and product category) partly stems from Craigslist's flagging policy.

The average number of effective ads posted per each hour is 26.2 and the standard deviation is 4. The average number of effective ads posted per product category is 157.3 with the standard deviation of 20.2. It is believed that the degree of variation observed in both distribution would not cause any significant bias in the collected dataset.

3) Collected emails and threads: The total number of emails received during the experiment was $19,204$ and the number of emails sent is $9,902$. Several emails in the same conversation are together referred to as a thread.

Among the total of $19,204$ emails received in our data collection $15,270$ were first responses. Among these first responses, our filter determined that $13,215$ represented scam-related activities, whereas the remaining include spams and fake PayPal payment emails and emails delivered from email service providers. As a result, our system attracted 9.6 scam trials (first scam responses) per ad.

From the $13,215$ scam-related first responses, our automated data collection engine sent $8,048$ first replies. As mentioned in Section III, presently we only send replies to emails that directly reply to our posts. There are $9,008$ out of $13,215$ first responses reply directly to our posts — by including the subject line that we used for our ads.

For $1,626$ of the threads, we received a second response from the scammer. Finally, we received 751 fake PayPal payment notifications emails and 885 bogus check fraud attempts. Note that we received multiple fake PayPal payment emails for some threads, and it was not always possible to tie a PayPal notification back to an email conversation thread, since for most fake PayPal notifications the source email address is different from those used in the email conversation.

B. Analysis of scammers' IP addresses

As described in III, we collected IP addresses of scammers by embedding an external link to a product image. We gathered IP addresses from web logs of the server that hosts product image files.

1) IP geolocation: In the experiment, we observed 965 IP addresses over 22 countries. The total number of accesses to the image hosting server from those IP addresses were $7,759$, and each IP address was observed 8 times on average. Figure 4 illustrates the IP geolocation of scammers who have accessed the embedded image links more than once. The

Overview	Duration of experiment Cities/areas Product categories	97 days (4/15/2013 - 7/19/2013) 20 (10 large and 10 small cities/areas) 4 (cell phone, computer, jewelry and auto parts)
Magnetic honeypot ads	Total number of ads	1,376
	Effective ads Flagged ads	629 747
Emails	Emails received Emails sent	19,204 9,902
Email threads	First scammer responses received First replies sent Second scam-related response received Fake PayPal payment emails (not threads) Bogus check payment threads	**13,215** 8,048 **1,626** **751** **885**

TABLE I. **Summary of experimental result**

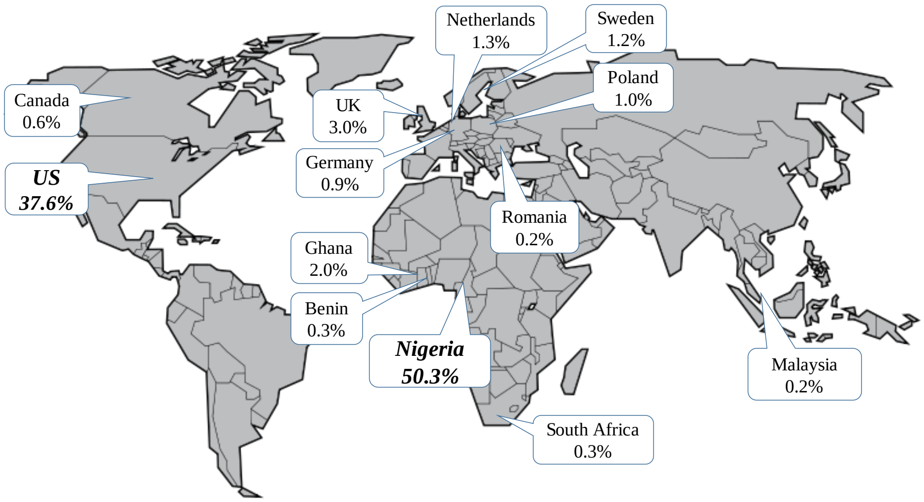


Fig. 4. **IP Geolocation of scammers.** For 965 IP addresses observed, 50.3% is are from Nigeria and 37.6% are from the U.S.

source `hostip.info`[4] was referenced to retrieve a geolocation information of each IP address.

Scammers' IP addresses were observed from all over the world but most of them were located in Nigeria and the U.S. In particular, 50.3% *of collected IP addresses were from Nigeria and* 37.6% *were from the U.S.* Note that this figure is plotted based on the number of unique IP addresses observed. It is also possible that some scammers could be using proxies, so the IP geo-location does not reflect their true location.

In Figure 5, the distribution of IP addresses over number of C and B class subnets is illustrated. We observed 413 class C subnets in total, and 40 of them take about half of whole IP addresses. Also, 10 out of 243 class B subnets account for about half of whole IP address. The result shows that small portion of subnets take major number of IP addresses observed, and it might imply the possibility of small number of scam factories dominating whole scam business.

2) IP blacklist: We cross-checked the collected IP addresses with a publicly available blacklist, *Project Honey Pot* [12] containing IP addresses of user-reported spam and and scam generators. The result is outlined in Table III. In particular, Project Honey Pot contains blacklisted IP addresses which were confirmed to be malicious; and graylisted ad-

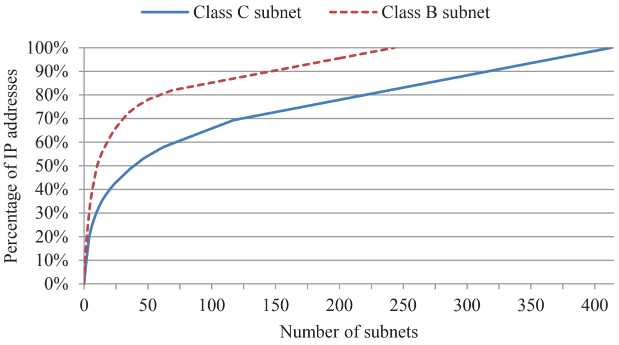


Fig. 5. **Cumulative distribution of IP addresses over number of subnets.** The total number of class C subnets is 413 and class B subnet is 243. Half of IP addresses observed belong to 40 class C subnets or 10 class B subnets.

IP addresses	Percentage
Not in black/graylist	43.9%
Blacklisted	40.4%
Graylisted	14.0%
Web crawler	1.7%

TABLE III. **IP addresses blacklisted by *Project Honey Pot* [12].**

[4] http://www.hostip.info

IP address	Country	# Times observed	Blacklisted?
41.211.193.XXX	Nigeria	298	-
41.203.67.XXX	Nigeria	241	Yes
41.203.67.XXX	Nigeria	204	Yes
41.203.67.XXX	Nigeria	160	Yes
41.211.198.XXX	Nigeria	93	-
41.206.15.XXX	Nigeria	89	Yes
41.184.21.XXX	Nigeria	89	Graylisted
41.206.15.XXX	Nigeria	88	Yes
41.211.201.XXX	Nigeria	85	-
41.206.15.XXX	Nigeria	79	Yes
41.206.15.XXX	Nigeria	77	Yes
41.220.68.XXX	Nigeria	73	Yes
41.203.67.XXX	Nigeria	71	Yes
41.206.15.XXX	Nigeria	68	Yes
41.206.15.XXX	Nigeria	64	Yes
65.55.255.XXX	The U.S.	62	- (Webcrawler)
41.206.15.XXX	Nigeria	60	-
41.206.15.XXX	Nigeria	58	Yes
41.206.15.XXX	Nigeria	57	Yes
41.206.15.XXX	Nigeria	56	Yes

TABLE IV. **Top 20 IP addresses by the number of times observed.** All but one are from Nigeria. The U.S. one is identified to be a crawler. 10 of the 20 IPs (in bold) belong to the same class C subnet.

dresses, which were detected but have not been confirmed to be malicious. From the IP addresses we collected, 40.4% are blacklisted by Project Honey Pot. 14% of IP addresses are in graylisted but not blacklisted. 1.7% of collected IP addresses are confirmed to belong to web crawlers such as *MSN search engines*.

About 43% out of the 965 IP addresses we found have not been blacklisted or graylisted by Project Honey Pot. Therefore, one contribution of our measurement study is to provide data to supplement existing IP blacklists.

Table IV shows the top 20 IP addresses by the number of access to the external link embedded in our replies. 19 of top 20 IP addresses are confirmed to be located in Nigeria. More interestingly, 10 of them belong to the same class C subnet, 41.206.15.0/24, which strongly implies that they are part of the same scam factory. One of top 20 IP addresses is located in the U.S. and confirmed to be Microsoft's web crawler. We suppose that the accesses from Microsoft were due to our use of Hotmail accounts. Other than one IP address of web crawler, 4 out of 19 IP addresses are not blacklisted yet.

C. Analysis of scammers' email accounts

1) Source, reply-to address discrepancy, and email account reuse: Throughout the experiment, we collected 4,433 email accounts used for first responses of 13,215 scam threads, indicating average reuse counts of 3 per an email account. The most frequently reused email account appeared in 101 threads. Figure 6 shows the distribution of email reuse counts. 2,410 email accounts, that is, 54.4% of total email accounts observed, were used in only one thread and about 10% were used in more than 6 threads. The majority of these single use only email addresses were initial inquiries about the product availability that never matriculated into further negotiations. However, many others were supporting emails used in the furtherance of the scam such as fake PayPal notifications, transportation agents, threats to contact the FBI (for when the product was not shipped) and similar emails. Some examples are posted in Appendix A.

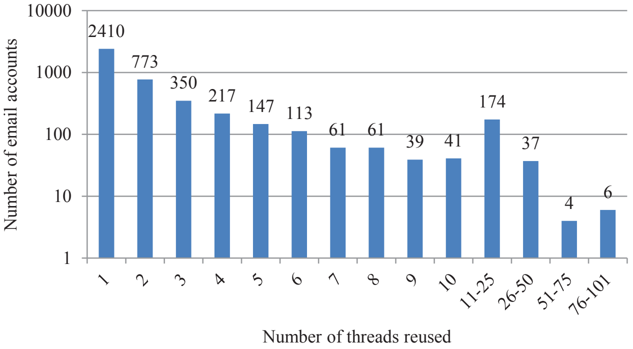


Fig. 6. **Distribution of email account reuse count.** For total of 4,433 email accounts, 221 are used for more than 10 threads, and one email account is used for 101 threads.

First responses	13,125
First responses with different source and reply-to addresses	10,826 (81.9%)
Second responses	1,626
Second responses with different source and reply-to addresses	316 (19.4%)

TABLE V. **Source and reply-to address discrepancy.** Source address and reply-to address are different in more than 80% of first responses, whereas the percentage is much lower for 2nd responses.

Number of source addresses mapped	Reply-to addresses
108	1
Over 50	5
Over 20	64
Over 10	141
Over 5	246

TABLE VI. **Number of source addresses mapped to a single reply-to addresses.**

We also observe that for 81.9% of the first responses received, the source email address is not the same as the reply-to address, and for 19.4% of the second responses received, the source email address is not the same as the reply-to address. This source and reply-to address discrepancy is shown in Table V. The percentage of discrepancy was much higher, 97.6% for first responses, within the top 10 groups. The operating procedures of top tier organizations must account for the increased quantity of emails sent and received, both for management and security, which is why they are more apt to split source and reply-to accounts.

Figure 7 shows that the set of source email addresses observed is much larger than the number of reply-to addresses observed. It is possible that the large pool of source addresses are disposable, and potentially automated, accounts that can be readily discarded and replaced as they are blacklisted. On the other hand the second, smaller tier of addresses are for more manageable monitoring and generally attempted to be kept "clean" for continued use over time. This intuition is supported by Table VI. We found a case that a single reply-to address was mapped to 108 source addresses. For total of 1,980 reply-to addresses, 141 were mapped to more than 10 source addresses.

Email provider	Percentage	Est. market share	IMAP/SMTP	Hide sender IP?	Price for 1000 PVAs
Gmail	**65.0%**	**25.0%**	**Yes**	**Yes**	**$90**
Microsoft	10.0%	20.3%	No (POP3/SMTP)	Yes	$5
AOL	4.9%	11.9%	Yes	No	$50
Yahoo	3.5%	42.8%	Yes ($20)	No	$15
Others	16.6%	–	–	–	–

TABLE VII. **Distribution of email service providers.**
Gmail is preferred by majority number of scammers, despite the highest PVA(Phone Verified Account) price, possibly since Gmail is the only webmail service provider which supports free IMAP/SMTP and hides email senders' IP addresses.

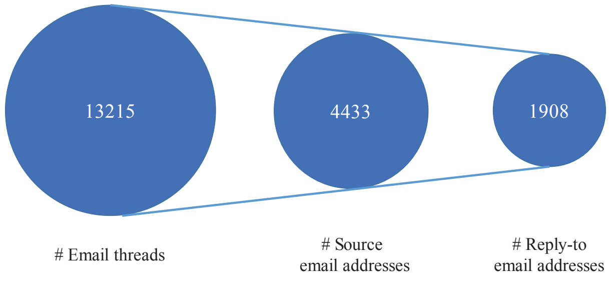


Fig. 7. **Number of email threads, source email addresses and reply-to email addresses.**

2) Email service provider: Table VII shows the proportion of each email provider the scammer uses, in comparison with the provider's estimated world-wide market share as reported by Geekwire [6]. We find that the top email provider used by scammers is Gmail, which accounted for 65% of the scammer email accounts observed, followed by Microsoft (Hotmail and Live), which accounted for 10% of the scammer email accounts observed. Interestingly, in terms of world-wide market share, Yahoo is placed first, with a market share of 42.8% [6]. However, Yahoo only accounted for 3.5% of the email addresses we observed.

It is interesting to correlate this with underground market prices of bulk email accounts. Many PVA (Phone Verified Account) sellers of black market sell Gmail accounts for a price higher than other emails such as Yahoo or Microsoft. For example, one PVA seller[5] sells 1000 Gmail accounts for $90, 1000 Yahoo accounts for $15, 1000 Hotmail accounts for $5 and 1000 AOL accounts for $50.

Despite the most expensive market price, scammers seem to prefer Gmail over other email services. The reason for this might be the fact that Gmail is the only email service provider who supports free IMAP/SMTP and also hides IP address of email senders among top 4 email providers. IMAP/SMTP is imperatively necessary for scammers, since they deal with massive amount of emails. Also, since scammers are working underground, they might want to hide their IP addresses not to expose themselves and possibly to avoid filtering by email service provider of the recipient. Scammers' aversion to Yahoo can also be clearly explained, since Yahoo charges $20 for SMTP/IMAP services. Although Microsoft supports free POP3/SMTP, those features were enabled recently in 2009.

Location	# Addresses	%
Nigeria	108	70%
USA	35	23%
Other	10	7%

TABLE VIII. **Shipping Addresses**

In this context, Gmail might be the reasonable solution for scammers.

3) Sample bad email addresses: We also observed interesting "clusters" of bad email addresses. For example, the following email addresses we observed are close variants of each other:

```
biglanre1@gmail.com
biglanre10@gmail.com
biglanre11@gmail.com
biglanre12@gmail.com
biglanre13@gmail.com
biglanre14@gmail.com
biglanre4@gmail.com
biglanre5@googlemail.com
```

This clearly shows that a scammer or a group of scammers create several email accounts to send out massive amount of scam emails.

D. Shipping Addresses and Phone Numbers

1) Shipping Addresses: 153 distinct shipping addresses were identified throughout the study by threads that progressed far enough so that shipment of the product was expected. As with IP addresses discussed in IV-B, the majority, 70%, of the shipping addresses were located in Nigeria with 23% and 7% located in the United States and other foreign countries respectively (Table VIII). Some shipping addresses had multiple names, assumed to be aliases, associated with them. In one circumstance, seven names were associated with a single Nigerian address. For the classification of the threads into groups, emails with the same shipping address were assessed as belonging to the same group. Additionally, some addresses were in close proximity to each other. For example, three different apartment numbers for the same street address in Nigeria were used as shipping addresses. In these circumstances, the threads were not assessed as belonging to the same group since being neighbors did not definitively indicate the occupants were part of the same organization.

2) Phone Numbers: 206 distinct phone numbers were identified during the study (Table IX). Most were given either as part of the initial inquiry or during the follow-on negotiation

[5] http://www.buybulkemailaccount.com/

Location	Service Type	Quantity
USA	VOIP	107
USA	Cellular	80
Nigeria	Unknown	12
Other Country	Unknown	3
Unknown	Unknown	4

TABLE IX. **Phone Numbers.** Unknown locations are due to missing digits or area codes.

emails, with only a few numbers withheld until the end of the purchase and then provided along with the shipping address. Diverging from the pattern seen with IP addresses and shipping addresses, the majority of the phone numbers, 91%, are registered within the United States and relatively balanced, but slightly in favor of, voice over internet protocol (VOIP) over cellular numbers. Of the 15 phone numbers identified as registered overseas, 12 were Nigerian, and all of these were associated with completed scam attempts and aligned with distinct Nigerian shipping addresses. Four phone numbers were either missing digits or area codes and therefore could not be categorized.

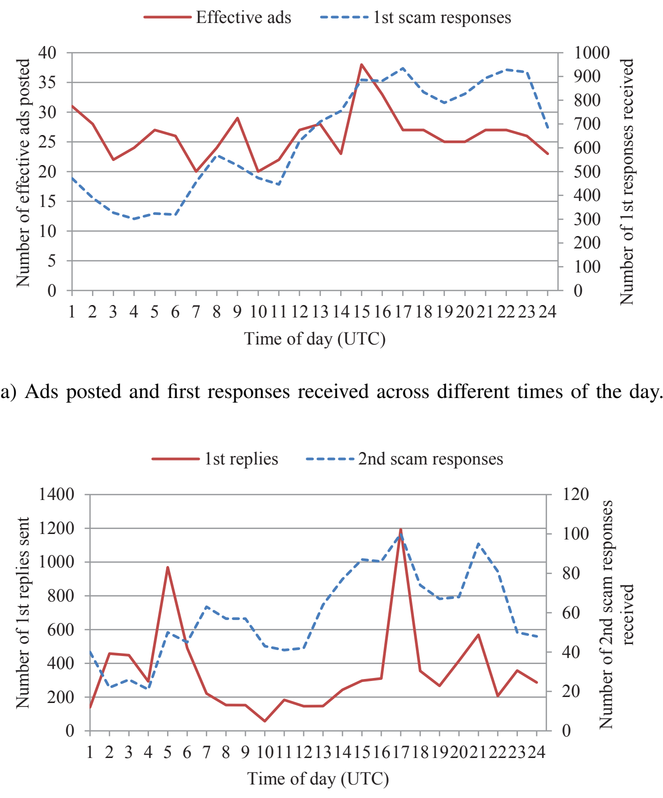
E. Scam Patterns

We report interesting patterns observed, including the distribution of scams received during various times of the day; response delay of the scammer; and what factors affect the scammer's response rate.

1) When do scammers work: The distribution of received time of first and second scam responses are illustrated in Figure 8. Figure 8a shows the distribution of posting time of effective advertisements and the received time of first scam responses. During the experiment, our automated data collection system posted advertisements evenly across the whole time of day. After removing the ads flagged by Craigslist, however, the number of effective ads varies slightly across different times of the day. Average number of effective advertisements at each hour is 26.2, the minimum number is 20 and the maximum number is 38.

Figure 8a and Figure 8b show that both the first and second scammer responses peak during 11AM to 6PM, and from 8PM to 10PM UTC (Coordinated Universal Time), This corresponds to the time period between 12PM to 7PM WAT (West Africa Time), which largely overlaps with working time in Nigeria. Moreover, the time period with the lowest scam responses is between 12AM and 6AM UTC, and it corresponds to 1AM to 7AM WAT. In addition, our first replies sent also peak during Nigeria's work hours, because our automated engine polls the emails every three hours and responds to the new emails. These observations support the result discussed in Section IV-B that the majority of the collected IP addresses are from Nigeria.

2) How fast do scammers respond: Figure 9 shows the distribution of scammers' response time. Only 6.5% of first scam emails were received within 6 hour and about 36% were received within 24 hours. The response time can be an indication of the level of automation of scam factories. We will discuss scam process automation later in IV-F.



(a) Ads posted and first responses received across different times of the day.

(b) First replies sent and second responses received across different times of the day. Our automated response engine sends replies within 3 hours upon arrival of a scammer email.

Fig. 8. **Received time of scam responses.** The peak time of both first and second responses (11AM to 6PM UTC) largely overlaps the business hours in Nigeria.

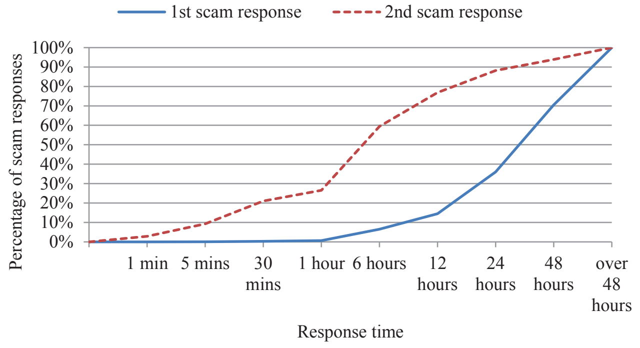


Fig. 9. **Response time of scam emails.** Only about 36% of first scam responses were received within 24 hours from ad posting. However, about 60% and 90% of second scam responses were received within 6 hours and 24 hours from our first replies, respectively.

On the other hand, we can observe much faster response time for second scam responses. 26.5% of second scam responses were received within an hour from the first reply of ours, and about 90% were received within 24 hours. This is likely due to the fact that our automated engine sends replies to scammers no later than 3 hours from receipt of their scam email. Figure 8b shows our first replies peak during the work hours in Nigeria. This explains why scammers respond more quickly to our first replies.

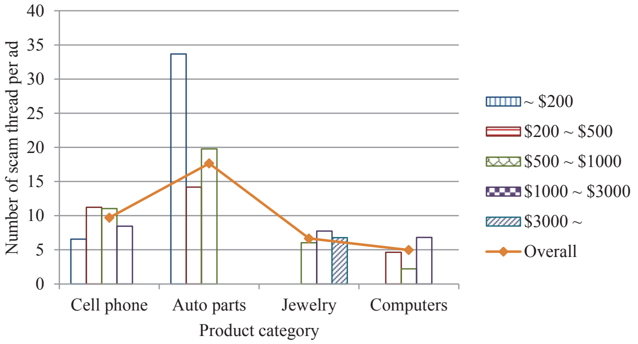


Fig. 10. **Number of first scam responses per effective ad.** Each of Auto parts ad attracted 17.6 first scam responses in average while each of computers ad attracted 5 first scam responses. For this plot, we used a subset of the scam threads which we were able to link to an ad.

3) Do product category and price affect scammers' response rate: As shown in Figure 10, each of our ads attracted 2.2 to 19.8 scam trials, depending on the category of the advertisement. The number of scam trials per *Auto parts* advertisement is 17.6 while the plot shows 9.7 for *Cell phone*, 6.7 for *Jewelry* and 5 for *Computers*. An interesting pattern observed here is that scammers seem to prefer a specific product category, *Auto parts*, over others. We ran one-way ANOVA test and confirmed that scam trials per auto parts ad were higher than other types of ads with P-value of 0.01. In addition, we confirmed that scam trials per cell phone ad were significantly larger than jewelry and computers. We are not able to state an explicit reason for the higher number of scam trials per ad for Auto part category in this paper. Reasoning for this observation would be left as a future work.

The number of advertisements posted by all Craigslist users is not a valid factor since we observed almost similar number of advertisements over 4 product categories used in our experiment. Also, price does not seem to be a valid factor since we are not able to find out any consistent pattern in 10 in terms of product price. We were not able to find out any correlation between number of scam trials and product price.

F. Level of Automation

One interesting question is whether the scam process is automated, and to what extent is it automated.

By combining various clues, such as inter-arrival time for the same email address and received email distribution across various times of the day, we can draw the conclusion that the scam process is automated to some extent, but not completely so. More details are provided in Table X and below.

1) Signs of automation: We observed clear signs of automation, including duplicate or templated responses observed in all stages of the scam process, outcome of broken scripts in subject lines, extremely short inter-arrival times for the same email address.

We now provide more explanation on the inter-arrival times for the same email address. Table XI shows some of the largest email bursts received. An email burst is a sequence of emails sent from the same email address within a very short time interval (no more than 15 seconds between emails). Table XI suggests that the largest bursts observed consists roughly 8 to 15 emails, with a mean interarrival time of 2.5 to 5 seconds. These bursts were not directed solely at Craigslist ads within a single city. One burst sent 15 email messages to multiple ads across 3 different cities. More interestingly, two of the largest bursts observed consist solely of second replies.

# Emails in Burst	Mean Interarrival Time	# Cities
15	5.2 sec	3
11	4.5 sec	3
11	5.6 sec	4
8	5.5 sec	3
7	4.3 sec	7
6	3.7 sec *	6
5	2.4 sec *	5

TABLE XI. **Example Interarrival Times for Burst Traffic from a Single Email Address.** All emails in the same burst have exactly the same content. (*: Second response emails)

We also observed many emails (including first response, second and later responses, as well as payment notifications) have exactly the same contents, or clearly use the same template to generate the content (Figure 13 in Appendix A). We also observed outcomes of broken scripts in first responses received. These demonstrate that the scammer are using some (semi-)automated tools to automate their response process, and more interestingly, their tools sometimes broke and generated email subject lines that are not human readable (Figure 15 in Appendix A).

2) Signs of manual labor: On the other hand, we also observe signs of manual labor. First, according to Figure 8, scam responses peak during working hours in Nigeria, which accounted for 50% of the IP addresses we observed in Figure 4. Second, we received Second and later scam responses containing curses — presumably the scammer was frustrated with us not shipping them the goods. Interestingly, we observed same curse emails occurring multiple times (Figure 17 in Appendix A). It is likely in this case that the angry scammer is copying and pasting the curse response. Also, in the latter stage of our data collection, some curse emails were received as a second response, before even reaching the payment stage — this could be a sign that the scammer actually started to detect our automated data collection.

The overall analysis of scam automation is outlined in Table X.

V. Classification

In order to discover how prevalently these scammers/organizations infected Craigslist and determine the scope of their operations, we classified the email messages into groups based on similarities within their attributes.

A. Conservative Classification Strategy

We first used a very conservative clustering strategy to classify scam activities observed into scammer groups. Specifically, if two scam threads shared *exactly the same email addresses, shipping address, or phone numbers*, they are grouped as the same scammer group. Email addresses whose prefix were 90% identical were individually reviewed along

Stage	Signs of automation	Signs of manual labor	Conclusion
Reading in Craigslist ads First scam response	short inter-arrival time of first response broken scripts in email subject duplicate/templated email contents	received emails peak during work hours	Both first and second responses are partially automated Scammers may need to manually run or attend to automated tools
Second and later scam response	short inter-arrival time of second response duplicate/templated responses	scammers' curse emails (Figure 15 in Appendix A), received emails peak during work hours	
Fake payment notification	duplicate/templated responses	Wrong email address/name in the notification	

TABLE X. **Analysis of scam automation.**

with other attributes such as email textual content and IP addresses so series such as the `biglanreXX@gmail.com` addresses noted earlier were also grouped together when multiple attributes showed similarities. In this way, we are highly confident that two scam threads belong to the same scammer group when we place them into the same cluster.

B. Top 10 Groups

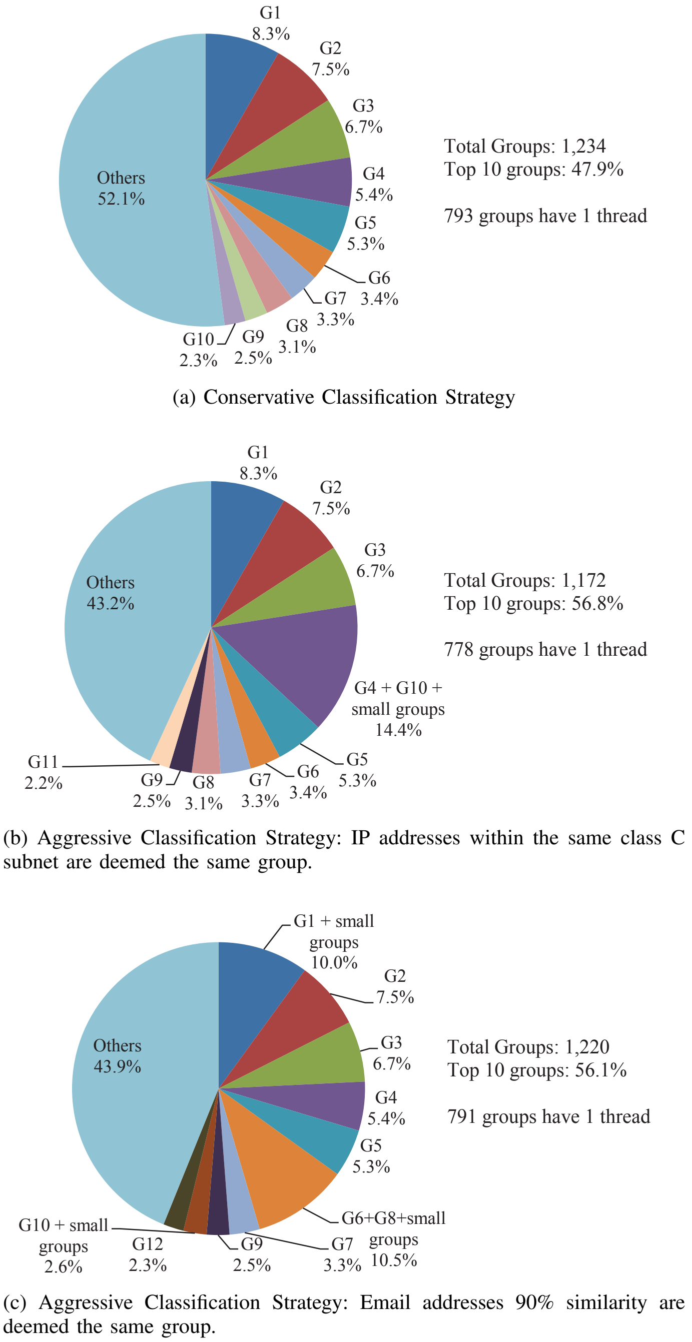
Based on our very conservative classification strategy, we found that *the top* 10 *groups accounted for* 48% *of all received scam threads* (see Figure 11a).

Further analysis of the top 10 groups showed that they operate over (almost) all cities where we posted ads (Table XII), and most of them throughout the entire duration of our data collection (Table XIII). Additionally, all groups responded to ads from all categories of products we advertised, *cell phone*, *computer*, *jewelry* and *auto parts*.

We give more detailed information about the top 10 groups below. Table XII lists details including the number of threads associated with the group, source and reply-to email addresses, associated shipping addresses and phone numbers for the top 10 groups by number of threads. Almost all of the top 10 groups had an extremely high ratio (ranging from 86.9% to 100.0%) of threads whose source email address is different from the reply-to address, Group 10 had a 86.9% ratio which was anomalous until it was discovered that a single email address was used as both the source and reply-to address in 11.1% of the threads.

Interestingly, only one of the top groups had shipping addresses associated with the group. Our assessment is that 9 groups are more sophisticated in their separation of initial inquiries and transition to scam attempts, successfully segregating the two in order to keep their clean accounts and personal information off blacklists. One group on the other hand appears unconcerned with this separation, working on a volume basis using the same email addresses and phone numbers for finalizing scam attempts that are used throughout the negotiating process.

Table XIII and Figure 12 show group activities over time. We make the following interesting observations. First, as mentioned earlier, all top 10 groups were active throughout the entire duration of the data collection. Second, Figure 12 shows that peak activities of a subset of the top 10 groups aligned with each other (e.g., the 5 plots on the left-hand side had aligned peaks and lulls). As mentioned later in Section V-C, some of the top 10 groups merged when we used slightly more aggressive grouping criteria — therefore, it is likely that in reality, a subset of the top 10 groups are actually the same big group. The aligned peaks and lulls as shown in Figure 12



(a) Conservative Classification Strategy

(b) Aggressive Classification Strategy: IP addresses within the same class C subnet are deemed the same group.

(c) Aggressive Classification Strategy: Email addresses 90% similarity are deemed the same group.

Fig. 11. **Group by Number of Threads. Small number of groups account for about half of scam threads.**

Group	# Threads	# Source Addresses	# Reply-To Addresses	% Source ≠ Reply-To	# Shipping Addresses	# Phone Numbers	# Cities	# Categories	Primary Category
1	1096	178	23	100.0%	0	0	18	4	Auto parts
2	993	270	64	98.7%	7	9	20	4	Balanced
3	885	313	48	95.8%	0	2	19	4	Jewelry
4	714	106	37	97.6%	0	0	20	4	Auto parts
5	700	52	11	98.6%	0	2	20	4	Balanced
6	449	182	30	98.0%	0	1	17	4	Auto parts
7	441	60	17	97.5%	0	1	20	4	Auto parts & Jewelry
8	416	103	10	100.0%	0	0	20	4	Auto parts
9	330	19	8	94.8%	0	0	19	4	Auto parts & Jewelry
10	306	71	23	86.9%*	0	0	20	4	Auto parts

TABLE XII. **Top 10 Groups.** Top 10 groups account for about 48% of emails threads. Scam emails of these groups were found in almost all of 20 cities and they covered 4 categories. They usually use a smaller number of reply-to addresses relative to the number of source addresses.

Group	First Email	Last Email	Duration
1	17 Apr 09:00	17 Jul 07:23	91 days
2	17 Apr 23:12	17 Jul 14:03	91 days
3	19 Apr 10:51	16 Jul 22:44	89 days
4	16 Apr 08:37	8 Jul 21:11	84 days
5	16 Apr 20:05	14 Jul 20:33	89 days
6	22 Apr 12:58	16 Jul 22:09	86 days
7	20 Apr 02:45	3 Jul 08:37	75 days
8	16 Apr 18:07	11 Jul 11:15	86 days
9	16 Apr 03:17	2 Jul 11:07	78 days
10	17 Apr 15:04	14 Jul 18:21	89 days

TABLE XIII. TOP 10 GROUP DURATIONS

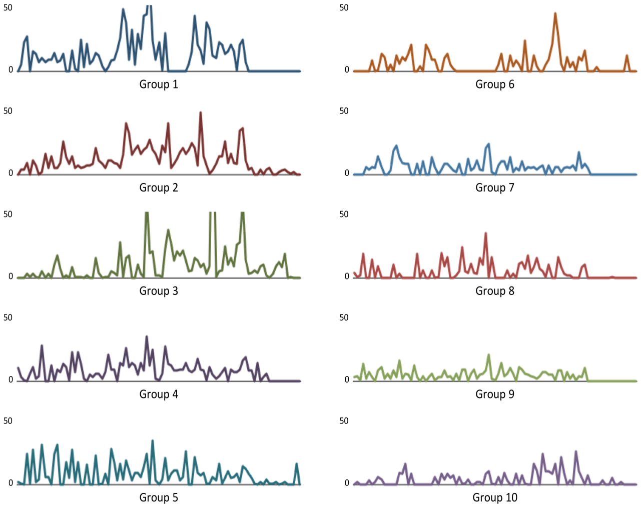


Fig. 12. **Emails Per Day - Top 10 Groups.** A subset of the top 10 groups show aligned peaks and lulls. Activities tapered off towards the end, partly due the lack of new posts and expiration of existing posts.

gives more evidence to support this hypothesis.

Finally, activities for most groups tapered off towards the end of the data collection, partly due to a combination of several reasons: the lack of new postings near the end of the research period, the expiration of postings created earlier, Craigslist's flagging of some of our postings, and additionally, some scammers may have started detecting our automated data collection.

C. *More Aggressive Grouping Strategy and Findings*

Our first grouping strategy in section V-A was extremely conservative. There are many other attributes that we could have used in the grouping, but chose not to. For example, for similar email addresses, while we manually inspected a subset of them and marked them as being the same group, such as the aforementioned `biglanreXX@gmail.com`; for others such as `madelineXX@gmail.com` and `alexandraXX@gmail.com` we conservatively chose not to mark them as the same, since Madeline and Alexandra are common English names.

We observe, however, that if one applied a slightly more aggressive grouping criterion, some of the top 10 groups would have merged. For example, Figure 11b shows the grouping results if we additionally merged IP addresses from the same class C subnet to the same group. In this case, the 4th and 10th largest groups from Figure 11a are merged, and is now the biggest group in Figure 11b. The remaining ordering of the top 10 are preserved.

Figure 11c shows the grouping results if we additionally merged email addresses 90% similar to the same group. The similarity between two email addresses A and B (containing only the substring before) is defined as:

$$\mathsf{sim}(A,B) := 1 - \frac{\mathsf{edit\_dist(A,B)}}{min(len(A), len(B))}$$

Intuitively, a 90% similarity between two email addresses means every 1 out of 10 characters may be different. In Figure 11c, the 6th and 8th largest groups from Figure 11a are merged, and is elevated to the largest group. The remaining ordering of the top 10 are preserved.

D. *Classification Summary*

Summarizing the above, our classification effort clearly indicates that *a small number of scam groups account for half or more of the total scam activities we observed.* These scam groups work across all cities, and were continuously observed throughout the data collection.

VI. LESSONS LEARNED

When we embarked on building our initial measurement collection infrastructure we did not know what type of replies we would be receiving to our magnetic honeypot advertisements. Based on our experiments and analysis we have discovered many potentially useful improvements to our infrastructure. In this section be will describe some of these challenges of our plans of deploying some potential improvements.

The first observation is that the scammers often do not include the correct subject in their replies to our advertisements. This causes our automated reply process to discard these messages, which is simple to overcome by replying to all messages received. The more challenging problem is that our analysis could not link these postings back to the specific advertisement the scammers were responding to because we were posting multiple advertisements with a single account. In addition, messages such as forged PayPal notifications were sent by the scammers that could not be linked due to different subjects from the original advertisement. In some cases we could manually link these messages to the correct advertisements. However, this could be solved by using an individual email address for each posting. We are implementing this by registering domain names, since creating large numbers of webmail accounts would violate their terms of service and would be labor intensive. However, from our initial results these custom domain names receive half the responses from scammers as compared to email addresses from well known webmail providers. This indicates that we might need to explore other methods of linking accounts using webmail accounts, since appear to be more attractive to scammers.

The second observation is that IP collection methods, such as embedding links to images in messages are effective at gathering information. We have improved this infrastructure to generate unique links that can be linked back to the advertisement. In the future we plan to improve this tracking by adding additional code to our server that will set cookies. These cookies will enable us to link multiple scam attempts to an individual if they use the same computer and do not clear their cookies.

VII. Discussion and Future Work

We have presented an in depth data-driven analysis of Nigerian scammers. This section will serve to provide a higher level view of our analysis to put it into context and discuss how our analysis might be used to deter these types of scams. In addition, we will describe future work that are planning to undertake that will improve our understand of these scams even further.

Larger Organizations. Our clustering results reveal that a large portion of the scam attempts are originating from a small set of groups that we can link together via their reuse of email addresses, shipping address, phone numbers, and similarity of the content in their messages. Our conservative estimate is that ten groups are responsible for about 50% of the scam attempts we received[6] This indicates that while this scam is highly prevalent that are only a relative small number of groups engaged in this activity. If these groups could be disrupted it would have a large impact on reducing the number of people targeted. As future work we plan to improve our measurement infrastructure to collect more information such as browser cookies that will enable us to more accurately cluster these groups. We will also work on improving our ability to estimate how many individuals are involved in this scam and the division of labor among the individuals within each group.

Locations of Scammers. We also find that all of major groups are based in Nigeria based on IP and shipping addresses. However, some of these groups use shipping addresses in both Nigeria and the U.S. This indicates that they might have some limited ability to receive packages and reship them to Nigeria. As future work, we plan to identify shipping addresses associated with the major groups and ship them items with GPS tracking units embedded into the device. This will give us more insight into how and where the stolen goods are resold.

Methods and Tools. Our analysis offers many clues about the level of automation and sophistication of the tools used by these scammers. We have found strong evidence of automated tools that are run manually or require manual attending. In addition, we find that these tools are used to automate both initial responses and in some cases follow-up responses. Also, some of these tools are able to crawl multiple geographic regions on Craigslist and parse the posting's subjects and contents to include in the reply. However, some the tools are fairly limited and include static text in the body of the reply or cannot parse subjects of listings. This relative lack of sophistication in these tools indicate it would be possible to incorporate the static messages into spam filters and at least force the scammers to develop more advanced tools. As future work we plan to design experiments focused on gaining more insight into the tools being used and their limitations. This might include crafting messages that ask questions targeted at better understand which messages from the scammers are automated.

Email Account Usage. Our analysis of email accounts used be the scammers, shows that they tend to use a large number of email address for the initial message that are quickly abandoned. However, they normally set the reply-to address in the initial message to a different email address that is reused often and longer lived. These longer lived email accounts offer a potentially better point of intervention, since it is conceivable these accounts can be blacklisted or banned before the scammer completes multiple email exchanges with the victim. In addition, as future work we will provide these email addresses to webmail providers and corporations such as PayPal to see if they have been used for other scams. This analysis might in turn help identify additional email accounts used by these groups.

Filtering Messages. In the course of our analysis we also identified many recurring themes in the content of the messages. These included the scammers claiming to be overseas military personnel to explain why they were located in Nigeria. Including religious content in their messages to gain the trust of their victim. Finally, they often used abusive language to coerce their victims into actually shipping the items. In addition, it would be feasible to exploit common linguistic features of scam email contents [4]. The combination of these patterns and the different reply-to address might be effectively used to improve the filtering of these messages. As future work will test this hypothesis by building improved filters that are more effective at detecting Nigerian scam messages.

[6] Note as we become less conservative with our clustering criteria some of the small group begin to merge into the larger groups and some of the larger groups begin to merge together.

VIII. Conclusion

In this paper we have presented a large scale empirical analysis of targeted Nigerian scams observed on Craigslist. From this we have learned valuable information on a variety of scam patterns such as scammers' working time and their response time to our ads and emails and discussed a degree of automation of scam process. Our analysis of IP addresses and shipping addresses indicates that the majority of scammers are located in Nigeria, but there is a smaller presence in the USA. We also found that around 10 groups account for almost half of scam attempts. Finally, we presented some higher level discussions based on our analysis and identify some potential points along this scam to intervene that might prove to be effective at deterring these scams.

Acknowledgement

We would like to thank Bobby Bhattacharjee and our anonymous reviewers, for their helpful feedback and guidance. We also thank Yali Li for offering crucial help with our data collection.

This work was supported by National Science Foundation grants CNS-1314857 and NSF-1237076, and Google Research Awards.

References

[1] Jim Buchanan and Alex J Grant. Investigating and prosecuting Nigerian fraud. *United States Attorneys' Bulletin*, 49(6):39–47, 2001.

[2] Marilyn A. Dyrud. I brought you a good news: An analysis of Nigerian 419 letters. In *Proceedings of the 2005 Association for Business Communication Annual Convention*, 2005.

[3] Kathleen Fearn-Banks. *Crisis communications: A casebook approach*. Routledge, 2006.

[4] Yanbin Gao and Gang Zhao. Knowledge-based Information Extraction: a case study of recognizing emails of Nigerian frauds. In *Natural Language Processing and Information Systems*, pages 161–172. Springer, 2005.

[5] Vaibhav Garg and Shirin Nilizadeh. Craigslist Scams and Community Composition: Investigating Online Fraud Victimization. In *International Workshop on Cyber Crime*. IEEE, 2013.

[6] GeekWire. http://www.geekwire.com/2011/stats-hotmail-top-worldwide-gmail-posts-big-gains/.

[7] Cormac Herley. Why do Nigerian Scammers say they are from Nigeria? In *WEIS*, 2012.

[8] Jelena Isacenkova, Olivier Thonnard, Andrei Costin, Davide Balzarotti, Aurelien Francillon, and France Eurecom. Inside the SCAM Jungle: A Closer Look at 419 Scam Email Operations. In *International Workshop on Cyber Crime (IWCC 2013)*, 2013.

[9] Chris Kanich, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M Voelker, Vern Paxson, and Stefan Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 3–14. ACM, 2008.

[10] Maria Konte, Nick Feamster, and Jaeyeon Jung. Dynamics of online scam hosting infrastructure. In *Passive and Active Network Measurement*, pages 219–228. Springer, 2009.

[11] Damon McCoy, Andreas Pitsillidis, Grant Jordan, Nicholas Weaver, Christian Kreibich, Brian Krebs, Geoffrey M. Voelker, Stefan Savage, and Kirill Levchenko. Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs. In *Proceedings of the 21st USENIX conference on Security symposium*, pages 1–1. USENIX Association, 2012.

[12] Project Honey Pot. http://www.projecthoneypot.org/.

[13] Aunshul Rege. What's love got to do with it? Exploring online dating scams and identity fraud. *International Journal of Cyber Criminology*, 3(2):494–512, 2009.

[14] Andrew Smith. Nigerian scam e-mails and the charms of capital. *Cultural Studies*, 23(1):27–47, 2009.

[15] Frank Stajano and Paul Wilson. Understanding scam victims: Seven principles for systems security. *Communications of the ACM*, 54(3):70–75, 2011.

[16] Brett Stone-Gross, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. The underground economy of spam: a botmaster's perspective of coordinating large-scale spam campaigns. In *Proceedings of the 4th USENIX conference on Large-scale exploits and emergent threats*, LEET'11, pages 4–4, Berkeley, CA, USA, 2011. USENIX Association.

[17] Brett Stone-Gross, Andy Moser, Christopher Kruegel, Engin Kirda, and Kevin Almeroth. FIRE: FInding Rogue nEtworks. In *Proceedings of the Annual Computer Security Applications Conference (ACSAC)*, Honolulu, HI, December 2009.

[18] Charles Tive. *419 scam: Exploits of the Nigerian con man*. iUniverse, 2006.

[19] Monica T Whitty and Tom Buchanan. The online romance scam: A serious cybercrime. *CyberPsychology, Behavior, and Social Networking*, 15(3):181–183, 2012.

Appendix

A. Example emails

Frequently Observed Emails

[Body:] `HELLO..IS THE ITEM HELLO, IS THE ITEM POSTED ON CL LISTED ABOVE STILL FOR SALE?? KINDLY GET BACK TO ME WITH THE LAST PRICE AND PRESENT CONDITION.THANKS`
# Times observed: `295`

[Body:] `HELLO, IS THE ITEM POSTED ON CL LISTED ABOVE STILL FOR SALE??`
# Times observed: `293`

[Body:] `Is it still available?`
# Times observed: `182`

[Body:] `Good day as i come across your listing on craigslist and i would like to know if its still for sale.`
# Times observed: `180`

Fig. 13. **Sample recurring emails.** Usually observed in first scam responses.

Belligerence/Threats

- [words omitted] Please respond to this mail before the penalty decision is taken against you. You are warned. We are waiting for your mail before we can credit your account and this is due to the large increase in the rate of the online scams recorded in the previous years. [words omitted]
- [words omitted] i think if i did not hear back from you within next 24hrs iwill have to contact FBI about your actions on Craigslist.org [words omitted]
- [words omitted] i will report you to paypal an FBI i give you 12h to get it ship [words omitted]
- Hey man what is going on i getting the FBI involve in this; is getting irritating [words omitted]

Fig. 14. **Sample emails with belligerent tones.** After making fake PayPal payment, scammers urge the victim to send the goods immediately. Note that some sentences are omitted for brevity or to remove personally identifiable information.

Broken Subject & Body

[Subject:] `<span class=i h data-id=0:00x0x_6FPl-zGRnuHX> </span>`

[Body:] Is your <span class="i h" data-id="0:00x0x_6FPlzGRnuHX"></span> still available for sale??
i will reply right away. Thanks
Sent from Devon’s IPhone

[Subject:] `<span class=i h data-id=3Fa3Le3I45Nd5I-c5Gcd624761cea879bf1558.jpg></span>`

[Body:] <html><head><META http-equiv="Content-Type" content="text/html;charset=utf-8"></head>
<body>still for sale?; feel free to email me at darrenamos69@gmail.com</body></html>

Fig. 15. **Sample broken subject and body lines.** Similar broken subjects were observed 3316 times in total. This observation implies scammers might be using some automated tools.

Email Bursts

[Body:] I’m interested in buying the posted item & your price Get back to me with your email welchcarrie619@gmail.com
Date/Times: 6/23/2013 3:04 - 4:47 PM
# Times observed: 36

[Body:] I’m interested in buying the posted item & your price Get back to me with your email sanjossmith@gmail.com
Date/Times: 6/29/2013 2:08 - 3:57 PM
# Times observed: 14

[Body:] I’m interested in buying the posted item & your price Get back to me with your email robert.waddick@gmail.com
Date/Times: 6/30/2013 3:56 - 6:04 PM
# Times observed: 19

Fig. 16. **Sample recurring emails bursts.** Bursts of emails are frequently observed. This observation also implies scammers might be using the automated tools.

Curses Indicating Manual Operation

- [curse word omitted] you stupid scammer. [words omitted]
- ARE YOU TRYING TO SCAM ME OR WHAT?
- WTF are you sending to me again ? i have already transfer the money into your PayPal Account and the money has already been deducted from my account;get the iphone 5 ship out via usps express mail and get back to me with the tracking number immediately you shipped. [words omitted]

Fig. 17. **Sample emails indicating manual operation.** In some cases, scammers detected us and sent this kind of curses in second responses.

Invoking God for Sincerity/Empathy

- [words omitted] I need you to be honest with the sale as I am a God fearing person. [words omitted]
- [words omitted] Note: I will be paying you extra money to cover the shipping cost through USPS EXPRESS MAIL. Also i wanted you to consider this sold to me and please remove the post from the craigslist site.Thank you and God Bless [words omitted]
- Do you still have this item for sale? GOD BLESS AMERICA.......................
- [words omitted] God bless as you do ship and i hope to do more business with you. [words omitted]

Fig. 18. **Sample emails invoking God.**

Capitalized Text

- LET ME KNOW IF THE ITEM STILL AVAILABLE FOR SALE.
- HELLO..IS THE ITEM HELLO, IS THE ITEM POSTED ON CL LISTED ABOVE STILL FOR SALE?? KINDLY GET BACK TO ME WITH THE LAST PRICE AND PRESENT CONDITION.THANKS

Fig. 19. **Sample emails with CAPITALIZED text.**

Recurring Themes in Emails

Military member unable to come view product

- [words omitted] My mode of payment would be in CERTIFIED CHECK and i will arrange for a local pick up as soon as you get the check; because that is the only convenient means for me and due to my work frame i can not be able to get there and i promise everything will go smoothly.I really wish to be there to check out the item but i don’t have chance cause am very busy person (US MARINE). And am already back to camp but i will get home very soon [words omitted]
- [words omitted] i have no problem with the amount as am a US marine i work for the United State Marine Corps (USMC) but am currently hospitalized so am on a treatment in New york [words omitted]
- [words omitted] Am willing to buy and am a serious buyer but am not around now so i won’t be able to come to have a look because am in camp now I’m a Marine(US MARINE).payment will be done by BANK CERTIFIED CHECK [words omitted]

Present for family member and buyer overseas

- [words omitted] i want you to know you are also in safe hands and i want you to assure me that i won’t be disappointed with it cos am getting it for my cousin the issue is that am not around i would have come and see it [words omitted]
- [words omitted] i wanted to buy this for my Cousin; but the issue is am currently out of state on a Contract Project .The contract is strictly no call due to the lack of reception in the area. [words omitted]
- [words omitted] im arranging it for my cousing birthday who live in OKLAHOMA USA.im off shore and Right now the only way i can make the payment is via paypal as i don’t have access to my bank account online and theres no way i can issue out a check or something here [words omitted]

Fig. 20. **Sample themes in emails.** Usually observed in second or later responses. Conversation leads to fraud attempt through fake PayPal payment or bogus check.