8-2020

Investigating and Validating Scam Triggers: A Case Study of a Craigslist Website

Hassan Mohamed Hirei
hhirei@go.stcloudstate.edu

Investigating and Validating Scam Triggers: A Case Study of a Craigslist Website

by

Hassan Mohamed Hirei

A Starred Paper

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Information Assurance

August, 2020

Starred Paper Committee:
Abdullah Abu Hussein, Chairperson
Lynn Collen
Balasubramanian Kasi

Abstract

The internet and digital infrastructure play an important role in our day-to-day live, and it has also a huge impact on the organizations and how we do business transactions every day. Online business is booming in this 21st century, and there are many online platforms that enable sellers and buyers to do online transactions collectively. People can sell and purchase products that include vehicles, clothes, and shoes from anywhere and anytime. Thus, the purpose of this study is to identify and validate scam triggers using Craigslist as a case study. Craigslist is one of the websites where people can post advertising to sell and buy personal belongings online. However, with the growing number of people buying and selling, new threats and scams are created daily. Private cars are among the most significant items sold and purchased over the craigslist website. In this regard, several scammers have been drawn by the large number of vehicles being traded over craigslist. Scammers also use this forum to cheat others and exploit the vulnerable. The study identified online scam triggers including Bad key words, dealers' posts as owners, personal email, multiple location, rogue picture and voice over IP to detect online scams that exists in craigslist. The study also found over 360 ads from craigslist based on our scam trigger. Finally, the study validated each and every one of the scam triggers and found 53.31% of our data is likelihood to be considered as a scam.

Keywords: Scam, Scam Triggers, Digital Infrastructure, Online business, Craigslist.

Acknowledgments

Let me start expressing gratitude toward my advisors through my starred paper research Prof. Dr. Abu Hussein, Prof. Dr. Lynn Collen, and Prof. Dr. Balasubramanian Kasi. This work would not have been managed without their bearing in any capacity whatsoever. I also need to extend my most enormous gratitude to my family for all the help you have demonstrated through this starred paper research.

Table of Contents

Chapter Page

List of Tables

Table Page

List of Figures

Chapter I: Introduction

Introduction

Today, online classified advertisements are a popular way to sell goods or services. The popularity of online classified ad websites such as Craigslist “www.craigslist.org”, Backpage “www.backpage.com”, Oodle “www.oodle.com”, and eBay Classifieds “www.ebayclassifieds.com” is continuing to increase. The World Wide Web provides a convenient and easily accessible medium for users to list and browse advertisements when compared to more traditional media such as newspapers and printed booklets. The widespread accessibility of the web has an unwanted effect of attracting online scammers who pose as genuine sellers by posting fake advertisements in an effort to defraud buyers (McCormick & Eberle, 2013).

Scammers have the ability to steal millions of dollars from unsuspecting users and threaten the reputation and utility of online ad services. There is no standard reporting of market or fraud statistics for online classified ads. Classified ad companies usually do not make public disclosures regarding revenue or fraud numbers. Victims may also not report occurrences of fraud because of embarrassment or uncertainty of where to make the report (National Consumers League, 2012).

To estimate the amount of fraud in online classifieds, we may consider the amount of revenue and popularity of such sites. Revenue from online classifieds needs to be differentiated from the amount of money that changes hands in online classified transactions. For example, the Internet Advertising Revenue Report (IAB) conducted by PriceWaterhouseCoopers lists online classified ads revenue at $2.6 billion for

2011(Price Waterhouse Coopers, 2012). It defines ad revenue as the fees advertisers pay to internet companies to list specific products or services. AIM Group's Classified Intelligence Reports projects the popular site, Craigslist, to have revenue of $126 million in 2012, an increase of 9.7 percent from the previous year (Zollman, 2012).

However, the vast majority of ads placed on classified sites are free. Considering that only a tiny percentage of ads are paid and that the person listing a paid advertisement anticipates an arrival or benefit, it is sensible to expect that the aggregate sum of cash traded through grouped promotion exchanges is more noteworthy than the site's income. Craigslist is the most well-known arranged advertisement site. As indicated by the web data administration, Alexa, it positions ninth in the U.S. furthermore, 42nd worldwide among all sites generally speak notoriety (Alexa, 2012). Craigslist's factsheet states that the website gets over 50 billion site hits and well more than 100 million ordered promotion postings every month (Craigslist, 2012). Other huge requested advertisement destinations that are not a long way behind incorporate eBay, Naspers "www.naspers.com", and Schibsted "www.schibsted.com". In certain territories, littler neighborhood grouped goals are progressively well known. With billions of notices put every year, including billions of dollars of exchanges, regardless of whether just a minuscule level of those promotions is a fake, it can defraud clients out of many dollars. With deals on the order of millions, these websites must monitor and attempt to detect potentially fraudulent activity. Their bottom-line is at stake, but even for those companies that do not charge for posting advertisements, their reputation can be compromised.

Craigslist is an online stage for ordered ads. Its effect on the neighborhood economy is non-immaterial—for instance, such notification decrease lodging rental opportunity rates (Kroft & Pope, 2014). Tenacious tricks limit the achievement of Craigslist. The chance of being misled could keep exchanges from being finished. In financial terms, the reasonable estimation of misfortune because of a trick can be considered a likely tax towards future transactions. Given Craigslist's beneficial outcomes for the economy, neighborhood governments should intend to limit this ordinary misfortune. Hence, nearby governments ought to either diminish the likelihood of exploitation or cutoff the extent of troubles. The undeniable arrangement is to diminish the number of assailants, for instance, through arraignment. Such deterrence is prohibitively expensive (Anderson et al., 2019), especially considering the limited resources of local law enforcement and associated public bodies. (The difficulties encountered in the application of consumer fraud statutes online are well-documented (*Policing Cybercrimes: Situating the Public Police in Networks of Security within Cyberspace*, 2010)).

Craigslist was begun in 1995 initially as a basic email appropriation arrange for promoting nearby occasions. The next year it was conveyed as a web administration. From that point forward, the administrations offered have extended and follow a nonexclusive worldview of publicizing the deal or willing acquisition of products and ventures. It is important that in 2004 eBay bought a 25% stake in the organization. These two administrations share the component of interfacing purchasers with dealers

however utilize totally various approaches. Craigslist, despite everything, makes most of its income from posting paid occupation notices (Lengle et al., 2009).

The Craigslist site has been divided by area, both county and city. This area division supports neighborhood trades, which expands security from extortion. The administration is additionally ordered into segments, for example, available to be purchased, work postings, personals, lodging, administrations, and so on. Inside these classifications, anybody can peruse and post new postings. Posting is substantial for 30 days before being expelled from the site. A post comprises data on the thing or administration to be sold or needed and some type of contact data for peruses to arrive at the banner.

Since the web administration was begun, it has been tormented by a progression of security issues, including mass phishing tricks and email gathering. The measure of help writing on tricks all through the site shows clear security issues with the framework. Nevertheless, there has been close to no change in the system in the past scarcely any years.

With the Internet's appearance, these scammers moved on the web, from the start in spam and simple messaging, and all the way to online grouped commercials and dating Websites. By targeting both businesses and individuals, these scams have resulted in financial losses of billions of dollars a year (Lengle et al., 2009), but also a psychological impact against their victims (Lengle et al., 2009). Today, the refinement of online scams keeps developing and improving as innovations permit scammers more scenes for access to casualties and improved capacities for adaptation of their scams.

Craigslist and other online sites have built up numerous defenses to filter through scam postings and shield users from misrepresentation, such a telephone check, boycotting IP locations, and observing for dubious substance. With that being said, the majority of those shields are centered around keeping scammers from posting fake data on the site; little effort has been made to shield authentic clients from accepting reactions from false purchasers. In 2013, a research estimation study was directed, concentrating on the phony payment scammer focusing on users on Craigslist (Lengle et al., 2009).

Advance charge extortion, all the more normally alluded to as Nigerian scammers or 419 scams, is a common type of online misrepresentation that not just purposes money related misfortune to people and organizations the same (T. Buchanan & Whitty, 2014), Yet, they can also carry passionate or mental harm to casualty clients (Jones & McCoy, 2014). An estimation of worldwide misfortunes to Nigerian scammers in 2005 is more than 3 billion dollars (Jones & McCoy, 2014). This kind of scam was initially generally untargeted and conveyed using email spam. Today, there are progressively modern focused on forms of this scam that are aimed at clients of classifieds, occupations, and dating Websites. Notwithstanding its commonness, the network's comprehension of focused online Nigerian scammers is as yet deficient. Numerous online sites, for example, Craigslist, screen through scam postings to ensure its genuine clients. For instance, Craigslist has numerous protections set up to forestall scam postings, for example, requiring telephone number check for a Craigslist record to keep scammers from enrolling huge quantities of Craigslist records and posting false ads, blocking dubious IP locations and accounts, and evacuating ads containing the

dubious substance (Jones & McCoy, 2014). In any case, little is done to protect clients from getting scam answers to their ads. Furthermore, email specialist organizations face an essentially additionally testing task when endeavoring to channel lower volume and target advance expense extortion spam instead of not so much focused on but rather more typical spam (e.g., pharmacy campaigns) (Jones & McCoy, 2014).

In summary, although scammers are very smart enough by using many ways to scam and post ads that looks like a real ad, there are many ways to know some of those ads posted by a scammer in any online platform such as craigslist, and yet there is no single study that has combined those identifications to help both buyers and sellers for their awareness. Therefore, this study aims to investigate as many as possible scam triggers, and validate those triggers by using ads posted in Craigslist to find an answer for the following main research question:

How to Investigate and Validate Scam Triggers: A Case Study of a Craigslist Website?

Problem Statement

In late 2018, Statista has reported that internet fraud is responsible for more than $100 billion of private companies and public losses (Infographic, n.d.). Be it blackmailing state or company facilities or scamming someone with an alleged inheritance from a forgotten relative, those scammers' tactics are pretty diverse. That implies nearly anybody can demonstrate an objective regardless of in the event that you are a small company, a moderately aged corporate worker, or a senior. As the Scam Tracker by the Better Business Bureau reports, web scam has soared as of late, with just about 46,000

documented tricks in 2017 in the U.S. alone and as of now more than 30,000 by mid-August of this current year.

According to United States Federal Trade Commission (FTC) report in 2019, Scammers like to get money by wire transfer for $423 million last year (Imposter Scams Top Complaints Made to FTC in 2018, 2019). With that direct wire transfer payment, there are also some other forms of payments, such as gift and reload or top-up cards – a 95% increase in dollars paid to scammers last year. Therefore, this study focuses on investigating and validating scam triggers in the context of the Craigslist website. This research aims to answer and address the following research questions:

1. What are the scam triggers to detect online scams on craigslist?
2. How to find and analyze Ads based on identified scam triggers?
3. What is the likelihood to be considered those ads as a scam?

Nature and Significance of the Problem

The advancement of the technology and web-based applications over the online scams gave the best result in the online advertisement such as Craigslist. In recent years, online scams have become a significant problem in online advertising, which has substantially affected the trust, beliefs, and encouragement of the customer on online marketing. Some of today's technology can find fraud or scam in web-based marketing and advertisement, yet that has not enough for the number of customers dealing with those technologies daily.

The Craigslist people reduce the scams as much as possible, but this can't be overcome so easily. If the users find that there is a posting which seems to be a scam or so they have to flag such kind of posting and so can the people of Craigslist can also look at those and delete them or drop, then into a bin category. To help their visitors, Craigslist offers advice in sections named "avoiding scams and fraud" and "Personal Safety Tips."

This study will first identify something we call scam triggers, which applies to Craigslist advertisements. Once those triggers are identified, we will take the second step, which is to validate those triggers by using advertisements posted in Craigslist today. Many or some of those scam trigger has already been identified, but no one has proved yet. Our study will be the first to list all possible scam triggers and take the additional step to verify or validate those triggers, and surely, this will be a significant help for both buyer and seller in Craigslist.

The objective of the Study

This study's main objective is to find out and list as much as scam triggers or identifiers as possible. Then to validate how likely those scam triggers are a scam by identifying ads from craigslist and then giving a close to attention couple days to those ads what happens after identifying them. Identifying ads from craigslist based on those scam triggers and then presenting the percentage of those ads are likely to be considered scams will be the entire aim or objective of this study.

Summary

In recent years we have become very reliant on the Internet for most of our daily activities; these activities include selling and buying items. People use many different platforms to sell and buy different goods from a house, cars, cloth, and shoes. This study will take the first step of identifying some scam triggers where someone that includes these in their ads the likelihood of it being a scam is high. The study will be used to validate those scam triggers in ads with the context of Craigslist. This study will not claim that those ads are a scam; instead, we say the likelihood that it's a scam is high if the following triggers are there.

Chapter II: Background and Review of Literature

Introduction

This chapter will extensively explain and bring all related background around online scams. This research will first look at the background related to the problem and then the background related to the proposed solutions regarding this problem.

Background Related to the Problem

Many people still use the Internet for at least part of their quest for accommodation. This eventually led to the posting of fraudulent rental ads by profit-driven scammers, widely known as "rental scams" (Park et al., 2017). Given the omnipresence of online rental scams, we still lack a strong understanding of the online rental scam environment, and the rental scammers use various tactics to trick and exploit their victims (Park et al., 2017). Although most attempts to address this issue are based on filtering the messages, this is just the visible part of a well-honed series of schemes and systems set up to collect money from their brands.

A thorough description of fraud and its systemic effects (message posting, email addresses, scammers position, support firms, automated devices, and payment methods) is also a critical first step towards recognizing possible chain vulnerabilities that can serve as successful defender choke points (Clayton et al., 2017). This recognition and the avoidance of the trend, in particular, has contributed to the identification of weak points to undermine certain domain-specific risks, such as payment processing of counterfeit apps and webspam in any online domain in general (Clayton et al., 2017).

Recent literature on cybercrime considers victimization as an individual behavioral problem (Clayton et al., 2017). Nevertheless, all criminal actions and subsequent victimization are a part of the makeup of the culture. Community disorganization theory, for example, dating back to the early 1900s, explores the concentration of crime in particular neighborhoods. It studied slum neighborhoods, where the level of illegal activity remained constant irrespective of the ethnic resident community (Garg & Niliadeh, 2013a). This hypothesis indicates that communities with more homogeneous populations are less vulnerable to crime Heterogeneity alleviates the collective effectiveness of informal social controls through unsupervised peer groups (Huang et al., 2015). A possible explanation for American cities would be that communities that are overwhelmingly white would be vulnerable to fewer scams on the Internet.

The second theory of criminology, the theory of daily behavior, considers crime as a result of incentives for offenders, availability of goals, and guardianship (J. Buchanan & Grant, 2001). Online transactions are more a daily practice for younger and more qualified individuals and thus correlate with higher rates a fraud (J. Buchanan & Grant, 2001).

At the same time, fraudsters are less likely to target those on the opposite ends of the educational continuum, i.e., those without a high school diploma and those with a college degree (Isacenkova et al., 2013).

Therefore, a lower percentage of high school graduates would represent a reduced number of scams on Craigslist. Around the same time, a higher proportion of

individuals with a bachelor's degree can also be associated negatively with the number of scam reports on the Craigslist. Income and poverty levels also include victimization (Johnson, 2014).

Offline, high-income areas will be more focused, as there would be a higher payoff for crime risk, such as robbery. Similarly, targeted online cities with a higher per capita income are likely. The town's gender makeup may also be important, as women are more at risk than men (Kanich et al., 2008). We are also less likely to fall in for scams. Around the same time, women perceive increased risk of online transactions and are less happy with eCommerce (Konte et al., 2009). Therefore, women are less likely to respond to automotive scams based on Craigslist (Levchenko et al., 2011). A theoretical possibility is that a higher percentage of women in the population would result in fewer scams being released because there is a reduced risk of victimization.

Craigslist auto scams are a typical instantiation of fraudulent advance payments. One potential scam in the automotive segment of Craigslist would involve a scammer posting lucrative automotive advertisements. The quoted price for these vehicles is irrationally small; setting an unrealistically low price may allow the scammer to filter out the unlikely gullible individuals (Herley, 2015).

The scammer will receive responses from potential victims, given the current signal of an irrationally low price. The scammers will then seek to justify these people with tales that would explain the reason for the unreasonably low price. For example, the seller may claim to be an international traveler who, on short notice, must inexplicably move back to his home country. Transactions are made using third party

agents like BidPay, Squaretrade, or PayPal. Victims advance the money without getting the car back to the scammers (Garg & Niliadeh, 2013a).

One of Craigslist's weaknesses is the lack of protection around the flagging system used to decide whether a post is useful or spam. As the program is designed to be open to the public and openly moderated, posts can be flagged by all. Post can be flagged as: miscategorized, banned, over post/spam, or Craigslist best. Craigslist bases the system on the premise that given an average over a collection of votes, you get a rough estimate of the actual status (good or spam) of a message. That presumption, however, depends on the assumption you can trust the voters (Garg & Niliadeh, 2013a). The flagging system can be misused to do one of two things. One can flag a post as best of Craigslist even though it is not. This option can help boost a post, resulting in postings that are not necessarily helpful, residing in the listings section's best. This removes a spot in the listings for a post, which is convenient, and it generates undeserved advertising for the poster. An advertising company could exploit this vulnerability. The company could post an add and flag themselves as best of Craigslist enough times that the system puts them in the 'best of' listings. The other misuse of the system is slightly more malicious. One can flag posts as spam even though they are not. This results in posts being removed from the system, which are legitimate.

This vulnerability could be exploited by any user who wishes to remove the competition from the system. Simply flagging any competitive posts enough times will have them removed, only leaving the user's post visible to visitors to the site. Keep in mind that advertising is one of the primary revenue sources for major companies like

Google and even Craigslist itself. The asset value of exploiting Craigslist for free advertising is higher than you might expect, and the flagging mechanism lends itself to these kinds of exploits (Garg & Niliadeh, 2013b).

Nonetheless, previous research focused on understanding the fundamentals of Nigerian scams targeting Craigslist and organizing the scammer groups (Park et al., 2014). Some studies have looked at the cultural challenges faced by the Nigerian youth, which has given rise to the prevalent scammer culture (Ojedokun & Eraye, 2012), (Aransiola & Asindemade, 2011).

Besides, multiple reports have centered on the Nigerian 419 scams and associated advance fee fraud methods across various channels (Smith, 2009) and other scam methods aimed at both social media and online auction sites (T. Buchanan & Whitty, 2014), (Rege, 2009). Herley's work (Herley, 2015) explains how scammers use their communications to target their efforts and weed out victims who are unlikely to fall for their scam attempts to identify potential scammers and weed out legitimate users of Craigslist. Several previous studies have looked at Smith's structure (Smith, 2009), and (J. Buchanan & Grant, 2001), and estimated advance fee fraud losses.

Literature Related to the Solution

It is notable that (Park et al., 2017) performed the first detailed empirical study of the online rental scam environment, as seen through the lens of Craigslist's rental section. A thorough analysis of these rental scam campaigns has been undertaken that helps us tackle concerns aimed at enhancing our understanding of the supporting infrastructure to identify alternative ways to disrupt this environment, such as: "What are the specific underlying scams?" “Where are the scammers, and what resources are they using?” "What are the current defenses effective?” "What payment forms do they use?". Finally, they summed up excellent achievements and observations as going down below. They were able to classify several big rental scam schemes on the Craigslist by creating some successful detection techniques. Besides, they expanded the automated conversation engine Scam-baiter (Levchenko et al., 2011) to automatically contact suspected rental scammers, which allowed them to understand what support infrastructure they used and how they monetized their articles. They found about 29 K scam listings across the 20 cities they were tracking, over a 141-day duration.

They also consider a complex range of methods used to monetize their established rental scam campaigns. Those involve attempts to trick customers into paying for credit reports and "bait-and-switch" rental ads. Six of the seven major scam campaigns found used credit cards as they analyzed the payment method used. Most campaigns have relied on businesses licensed in the U.S. to collect payments. They have noticed that Craigslist's filtering systems are actually eliminating less than half of the payment methods. They also reveal new schemes and technology not found in

previous research (J. Buchanan & Grant, 2001), (Johnson, 2014) and (Park et al., 2014).

This disparity illustrates the need to consider a broader scam environment and indicates possible bottlenecks in regulatory and payment layers for other rental scam monetizing strategies. For example, regulatory authorities in the United States, such as the Federal Trade Commission (FTC), may investigate such companies and impose fines on their misleading advertising practices. Another possible way to demonetize these businesses might be to alert credit cardholder groups, such as Visa or MasterCard, to the misleading billing and refund policies of those retailers.

Most of the previous related research focuses on the use of content-based features to detect general webspam. (Fetterly et al., 2004) were among the first groups to find material for the identification of spam pages using statistical analysis of two datasets: DS1 contains 150 million URLs, and DS2 encompasses 409 million HTML pages. They also noticed that many spammers use templates to create spam pages automatically. These pages would then have exactly the same number of words, while the specific words can vary from one page to another. Fifty-five percent were spam in a survey of 200 pages taken from hosts with at least ten pages and without any difference in the word count.

The hypothesis is focused on the evolution of content on web sites. According to (Fetterly et al., 2004), for a week, 65 percent of all pages remain the same, and just 0.8 percent of all pages shift fully. Web servers of some spammers produce a response to any HTTP request without using the application's actual URL. The spam pages change

more often than in HTTP requests, and they are independent of the URL. In a survey of 106 server pages with all pages completely shifting in a week, 97.2 percent were spam. Another excellent research introduced a detailed taxonomy of webspam strategies including content-based rating algorithms.

They also give some suggestions for countermeasures to webspam. They also reported on a sample of over 105 million web pages accessed via crawler MSN Search. 86.2 percent are non-spam, and 13.6 percent are spam, in a survey of 17,168 pages taken from English websites. They extracted more than 20 features from the content of pages, which can provide details to differentiate between a spam page and the usual ones.

One way of enhancing the efficiency of detection is to combine content-based features with other features such as link-based features (Abernethy et al., 2008). In (Abernethy et al., 2008). Using the WEBSPAM-UK2006 dataset (Abernethy et al., 2008) to collect labeled web pages from .uk domains crawled in May 2006.

They added new features to the set proposed in (Ntoulas et al., 2006) , which could help separate spam pages from non-spam pages based on their WEBSPAM-UK2006 analysis. They also introduced link-based features and then used them to build the classifier along with content-based features.

Build a rating structure based on the PageRank algorithm. We build a web graph in which we measure two types of scores for each node, positive score for the node's authority, and the negative score for the likelihood of spam for the node. Content-based functions are used to change the scores for the node. Pages with high score negative

are more likely to be spam. (Fetterly et al., 2004) suggested a further content-based strategy for detecting spam pages using a shingling method to examine whether a page contains several common phrases copied from other web pages. They produced a fingerprint with each document and then attempted to compare the fingerprints of two separate documents to see whether they are duplicates or close duplicates. We found out that if a page has a large fraction of repeated sentences, it is likely to be spam. Use the fingerprints of papers, Urvoy et al. have attempted to identify the resemblance of secret styles between the pages (Urvoy et al., 2008).

They aimed at identifying pages that are created automatically using templates. Consequently, if they have a sample of a spam page, they will find other spam pages that are created using the same method even though they are typically distinct. One approach to identifying spam using content (Benczúr et al., 2006) includes using language models (Jay M. Ponte and W. Bruce Croft, 2019).

Its essence, a language model, is a distribution of probabilities over terms derived from a list, in this case. The language model approach's basic concept is that there should be a local connection between two related pages. Otherwise, the relationship can indicate that spam is occurring. Calculating the Kullback-Leibler divergence (KLD) value, calculating the difference in the language models, i.e., the probability distributions of the two articles, will define this relationship.

In (Martinez-Romo & Araujo, 2009), KLD used to calculate the difference between source-page text and target-page text. Derived from various content elements of the source and target pages, KLD introduced new features that, together with other

content-based and link-based features, help improve classifiers' performance. A change to the Latent Dirichlet Allocation (Bíró et al., 2008) is used in (Bíró et al., 2008) to build sets of topics for spam and non-spam websites using their word bag. Then they combine collections of those subjects. A new site is marked as spam if it has the probability of spam subject being total above a threshold (Piskorski et al., n.d.) also tried to extract the linguistic features for spam identification, based on text.

Its preliminary results are very promising ratio, brand, manufacturing year, etc. Some of those characteristics are defined using external resources. For instance, they extracted various car features from the advertising post, including year, make, model, etc. to capture whether the asking price for a car is fair. We then use external tools, e.g., Kelley Blue Book (KBB) or Edmunds4, to obtain an average price for the vehicle and equate it with the price asked. Or we can use Yahoo! Position Finder5 to estimate the distance from the seller to post location to see if the post is nearby. The next segment provides detailed explanations of certain functions.

Finally, it provided a feature vector for each article, and they transformed the problem of spam detection into a classification problem. Several well-established tools and techniques can then be used to solve it. Like some previous works detecting spam content on the site, a decision tree classifier was applied to classify the post into spam or non-spam.

Existing fraud or spam detection approaches include group monitoring, proprietary computer systems, Geo IP applications, and tools to prevent commercial fraud. Many users of the site flag offender listings during group moderation. It is either

deleted or submitted to the site's moderators for review if an ad earns many votes. Reduction in the Group has the advantage of being easy to enforce. It crowd’s sources of identification and eradication of fraud by relying on other human users' judgment and intelligence.

Society will likely adapt and accept new types of fraud when they are uncovered. Some disadvantages are depending on Group moderation. Once they can be flagged, consumers must be exposed to fake ads. Individual users may not have access to all the information. IP addresses and other data based on computers, such as personal account information are unlikely to be available.

Although human intuition and judgment may be helpful, users are not experts in recognizing fraud and may not be the best to detect it. Users can often, for some reason, exploit the program by flagging content, including rivalry or retaliation against other users. Websites also use their own patented fraud detection schemes. Such systems have the advantage of being automated and may integrate information unique to the domain. Proprietary systems can include a set of rules designed by hand. Every rule considers an ad attribute and either increase or decrease its value.

Every rule scores new listings, and then its value is compared to a given threshold. When the ad is above the threshold, then the suspect is labeled. Handcrafted rules depend on a domain expert to have the experience to establish rules that differentiate effectively between fraudulent and regular listings. Those rules and best weights can be difficult to design manually. Unless the rules effectively stop fake advertising, scammers are likely to adapt and modify their tactics, with the result that the

ruleset needs to be revised. Geo IP services like MaxMind (MaxMind 2012) have converted an IP address of a user into a geographic location.

This position can be likened to the commercial location. Unless the IP address matches a particular venue, this may suggest that the advertising is fraudulent. Although the Geo IP services are useful, more information is available, and fraud detection could be required. Furthermore, posting advertisements when out of town is not unusual for users because their position does not suit the listing. Tran and. Al. Suggested a system for identifying spam in classified ads online (Tran et al., 2011).

They received several months of ads that were posted on Craigslist. Volunteers then produced a training collection (McCormick & Eberle, 2013) from a limited selection of such advertisements by marking the ads as spam or not spam. A decision tree classifier was equipped to identify spam instances with features extracted from the advertising. With their classification system, they demonstrated substantial improvement over conventional Web spam detection methods. Although the authors give an example of a fake advertisement, they don't give a good explanation of the spam-fraud differences. Fixing a distinction between fraud and spam is critical.

Various motives and motivations for fraud and spam, and thus different signatures, may exist. Tools used to detect spam cannot be sufficient to detect fraud and vice versa. Spam is an unsolicited advertisement that provides flood services, such as classifieds. Spam is commercial and multinational, as opposed to classifieds posted by individuals to sell products or services to other local individuals. Spam also contains questionable items such as knockoffs, inexpensive prescription drugs, or schemes for

getting-rich-quick. Any effort can be made to mask the purpose of spam ads. The aim is to encourage consumers to pay for the advertised goods and maybe to sell actual products. The use of fraudulent advertising, by comparison, is to mislead the consumer by pretending to be a regular listing.

We seek to blend in with the hosting community by posting identical products and services to other already available ones. There is typically no specific commodity. The entire ad and history were designed to draw users into scams. The seller can, for example, claim to be out of town and request funds transferred electronically. The seller disappears after collecting the money, and the item being sold is nowhere to be found. Tran et. Al. 's trial, the volunteer training data showed that 17 percent of the advertisements were spam. Whether such a significant percentage of all advertisements are fake seems impossible. It might be concluded that the branded spam contained a mixture of spam and fraud without a clear distinction. The volunteers used to mark the training data are possibly not experts in detecting fraud, similar to group moderation. This is likely that even the most prominent instances of fraud have been identified.

Since spam may not be hidden and only simple instances of fraud might have been found, it is probably easier to train a classification model to detect these cases. The method (McCormick & Eberle, 2013) analyzed data generated by a classified advertising website. The website operators have also issued a list of advertisements previously found to be fake. They claim that this will provide us with the most reliable data possible as the site operators are likely to have the best expertise in labeling fraud. Furthermore, because the information is generated from an internal source, features

which are not made public can be derived from the data. Such characteristics provide such things as the user's IP address and date of accession.

They used the data to create a model of classification. This model may be used to detect fraudulent ads when displayed, as in the case of group moderation, which prevents users from being exposed to deceptive ads. Since information about the user doesn't need to be shared and monitored by a third party, there are fewer questions regarding user privacy recognition.

The ad archive also contains user account information and user activity reports. Private user information has been anonymized in compliance with the website's privacy policy to be able to interact with the database details. Therefore, before obtaining the data, names, street addresses, email addresses, and other personal accounts, information was deleted. -- ad record links to a user record that contains a unique integer identifier, the user's date of registration, location, state, and zip code. The site also holds a list of login information for the users. Every time a user logs in a record that contains login date, session ID, IP address, user agent, breadcrumb, and geo IP data is generated. The breadcrumb is a randomly generated 128-bit identifier stored in a cookie on the user's computer.

Their identifier aim is to uniquely identify user behavior through multiple sessions and even various user accounts. Geo IP data is extracted and stored from a web service, which provides additional information on an IP address with each login record (McCormick & Eberle, 2013). This geo IP data includes the country code associated with the IP address, area, city, postal code, latitude, longitude, and ISP.

The company also presented a table of announcements previously classified as fake using their new detection methods. A breakdown of the number of ads received appears although fraud accounts for less than 1 percent of the ads received, it still presents a major problem in day-to-day operations and poses a risk to website users. The difficulty lies in detecting such fake advertisements by locating the proverbial needle in the haystack.

A unit of analysis is restricted to a city performed by (Garg & Niliadeh, 2013a). The U.S. classifieds are listed under 413 cities within Craigslist. Chicago, Cleveland, Dallas, Denver, Detroit, Honolulu, Houston, Kansas City, Las Vegas, Los Angeles, Miami, Minneapolis, Nashville, New York, Orange County, Philadelphia, Phoenix, Portland, Raleigh, Sacramento, San Diego, Seattle, San Francisco, St. Louis, Tampa, and Washington DC.

Interestingly, they did not presume that the scammers are from the same jurisdictional area as the city for posting their ads. They concentrated on Craigslist's cars + trucks site, which publishes auto classifieds in two categories: by-owner and by-dealer. They (Garg & Niliadeh, 2013a) looked at the by-owner classifieds and left the by-dealer section inquiry to future research. Their collection of data began on 11/19/2010 before the Thanksgiving break and lasted for 60 days. They were using crawlers to capture previously posted advertisements. Therefore, for 3.5 months, the data collection constitutes ads.

Craigslist has a user-flagging program where unauthorized and unacceptable posts are detected. When a certain number of users flag an article, the flag is replaced

for removal message, and it is eventually deleted after a few days. We figured highlighted ads were scam advertising and spam advertising. An advertisement can, however, be flagged for other purposes. For example, the ad can be put in the wrong Craigslist section. Its data is noisy, therefore.

They thought the noise is distributed evenly around the dataset. Remember that flagged data doesn't represent all scams, because specific scams may not have been flagged for removal by the requisite number. We presume, however, that their data reflects the relative distribution of Craigslist scams. The total number of single ads detected was 2,424,092, of which 42,185 were flagged. They categorized the flagged ads based on the city they were posted after collecting data and then extracted the total number of ads and flagged ads per city (Garg & Niliadeh, 2013a).
Therefore, they extracted two dependent variables: 1) the total number of flagged ads, and 2) the percentage of flagged ads. Of the independent variables, they had two primary sources: 1) the U.S. Office of the Census and 2) FBI Crime Statistics.

To understand the extent of system instability (Lengle et al., 2009). They also developed a tool on Craigslist for use in auto flagging messages. They decided to manipulate the program to flag posts unfairly to get them deleted automatically. As described earlier, no external anti-spam techniques are utilized by the flagging system, and anyone including a computer may simply perform an HTTP post-operation to the flagging type. This method needs two inputs: the post ID, the unique numeric identifier attached to each of the system's posts, and a flag code. -- flag category has its numerical code value, i.e., flagging as spam has a 15 flag code and flagging as best

Craigslist has a nine code. Upon submitting the form to the server, the response is a web page containing the words "Thanks for flagging."

A program that checks this pattern in answer to ensure the flag has been submitted. Their final tool created was based on the Craigslist, as mentioned above, crawler. The software takes a region and category of interest as its input and a search string entered by a user (Lengle et al., 2009). Their software starts to crawl the web, and when it reaches a post with text matching the input string, it submits a flag form for a spam post using the posting ID and the flag code. The program performance is a list of flagged posts as well as post-flagged statistics vs. crawled posts. Back end that deletes posts only under such circumstances.

Such requirements can depend on the posting age and the user flagging type, logged in versus anonymous user. In this case, a post flagged 15 times, each from a different IP is much more likely to be removed automatically than a post that is not. Consequently, our approach undermines the system's credibility but not to the extent that a malicious user would have hoped for. Using an unknown algorithm to prevent malicious behavior violates one of the principles of designing safe systems. To preserve the protection, a system should not rely on design secrecy. With the ability to further analyze and check to determine the exact criteria needed to automatically remove a message, our auto flagging system will become a very dangerous tool capable of removing any desired article.

Although previous work concentrated exclusively on collecting online analytical data collected via automated email encounters with scammers (Jones & McCoy, 2014),

additional data elements collected via manual online and offline encounters were introduced, mainly by examining the fake checks sent scammers and their methods of transport. Another series of works have focused on the role of money mule in the extraction and anonymous exfiltration of money from the victim into the scammer (Florencio & Herley, 2015), (Aston et al., 2009). Moore's work (Moore et al., 2009) describes how "the mule is directly responsible for the funds that have already been sent."

These works concentrate mainly on the role of mules in phishing scams, while we are investigating the related role of money mules in the online advertiser scam culture, including the role that the victim plays within the money mule chain.

The study by (Nilizadeh et al., 2011) focused on collecting empirical data to enable a data-driven analysis that does not rely on self-reported statistics. (Isacenkova et al., 2013) identified a thousand scam groups using a multidimensional clustering technique from an existing scam email dataset. Their research also argued that the email addresses and phone numbers of scammers are In Possession of a Toyota Sienna: Auto Scam Prevalence and Recognition (Nilizadeh et al., 2011).

Craigslist ads are accessed every month by millions of Internet users, making this an enticing target for fraudsters and miscreants. Unsurprisingly, a "cesspool of violence," has also been branded. In this article, we take a first look at the Craigslist auto scam. Focusing on the U.S. market, we find that scammers manipulate the fact that Craigslist ads are free to publish. For certain cities, they post a large number of advertisements for the same vehicle in a short period of time, either manually or by

exploiting the ease of automated ad posting software is available. Ironically, scams also advertise and list at enticing rates, fairly new automobiles of famous makes. We use special characters extensively for attracting attention and randomizing the ad body to avoid automatic detection.

Thankfully their analysis finds several features that differentiate between a scam and legitimate advertising. Using these features, we show that an SVM-based classifier can discern 99 percent accuracy between a scam and trustworthy advertising (Nilizadeh et al., 2011). Goa et al. examined the use of ontology-based information engineering for text mining of Nigerian scam emails. In comparison to previous research, their study centered on 1) understanding the nature and tactics in great detail, and 2) defining the organization of larger-scale groups of scammers involved in trying to defraud people advertising products for sale on Craigslist.

A wide body of recent research has set out to perform analytical analyses to clarify the nature and economic underpinnings of different kinds of cybercrime. Most of this research centered on spam communications, illegal online pharmacies, and mapping out the framework for scam hosting. Their work builds on this but is strongly focused on the issue of the Nigerian scam. The first large-scale systematic assessment study of 419 scams was performed. It gave us insights into how these scams are being orchestrated and how they could be better deterred.

Advance fee fraud, more commonly called Nigerian scams or 419 scams, is a prevalent form of online fraud that not only causes financial damage to both individuals and companies but can also cause emotional or psychological harm to victim users.

An estimate of global losses to Nigerian scams in 2005 is over $3 billion. Originally this scam was mostly untargeted and delivered via spam email. Today, however, there are more advanced targeted variations of this scam intended for users of classifieds, work, and dating sites. Despite its prevalence, there is still a lack of awareness by the community about targeted Nigerian online scams. Most websites online, including Craigslist, weed out spam posts to protect its legitimate users.

For example, Craigslist has many protections in place to avoid spam messages, such as requiring verification of telephone number for a Craigslist account to avoid scammers from registering large numbers of Craigslist accounts and posting fake ads, blocking suspicious IP addresses and accounts, and deleting ads that contain suspicious material.

Yet nothing is done to prevent users from getting scam responses to their ads. Furthermore, email service providers face a substantially more difficult challenge when attempting to handle the lower volume and target advance fee spam fraud rather than less targeted and more prevalent spam based on Nigerian scams on Craigslist, one of the most popular online advertising websites whose monthly visitors in the U.S. alone are over 60 million. They provided a detailed analysis of indicators of these scam activities.

Via this analysis of measurement, they aimed to better understand Nigerian scams' underground economy and search for strong points of intervention. In particular, they are trying to answer concerns such as: "Where are the scammers? "How do factories of scam work? "What features can we use to differentiate between a legitimate

email and a scam email? "They posted magnetic honeypot ads to better understand Nigerian scams on Craigslist-designed to draw scammers but repel authorized users.

They received and answered scam emails from our ads and analyzed the responses. They developed an automated data collection system for quantitative scam analysis, which posts advertisements, collects scam emails, and communicates with scammers by sending out responses to the scam emails received.

They also gathered scammer's IP addresses to accurately validate the scammer's geolocation. They have done various analyzes of the massively collected data set to better understand how scammers function. They also observed cluster scammers into groups based on a few key factors: email addresses, shipping addresses, phone numbers, and payload emails. Our research shows that these forms of Nigerian scams are prevalent as our magnetic honeypot advertisements earned 9:6 scam responses on average. Our study's most enlightening finding is that about 50 percent of the scam attempts observed can be related to the top 10 categories.

Such groups promoted advertising distributed through many types of Craigslist products and geographic regions. However, their research found that many of the initial scam messages are coded and arrive from several quickly discarded email addresses. Nevertheless, several of these initial messages included a different answer-to-address to a smaller collection of email addresses with longer life. They also found that 23 percent of shipping addresses were in the United States, while most IP addresses and shipping addresses were in Nigeria. It suggests that either accomplices or re-shipment activities were likely to have been used. Their review of the messages' content indicates

some instances of terms such as God, military personnel from abroad, and capital letters that may help filter such messages. They identified multiple possible intervention points from their study.

Their review of the message content, in particular, suggests that the filtering of messages may be enhanced by checking for variations of such patterns such as a response-to-address that does not match the address of the sender, the use of such odd phrases, and the detection and blacklisting of these more secure and long-lived secondary accounts. The mailing addresses may also be the starting point for inquiries into law enforcement. On the same lines, the fact that only ten groups of scammers accounted for almost half of the scams, we got suggests that these groups may be targeted and disrupted, significantly reducing this scam incidence.

Literature Related Gap Analysis

The table below is the essential source used to analyze the gap in our research area. The papers listed in the table are the most up-to-date papers published and the nearest papers relating to our field or study area. The table presents each article, and the paper focused on the problem and its weaknesses—shortcoming what helped our work narrow down our objective and ensured pickup from there.

Table 2.1

Gap Analysis Summary

Paper	Problem	Shortcoming

Vaibhav Garg and Shirin Nilizadeh, "Craigslist Scams and Community Composition: Investigating Online Fraud Victimization", 2013 IEEE Security and Privacy Workshops, DOI 10.1109/SPW.2013.21	• Examine the prevalence of Craigslist-based (automobile) scams across 30 American cities.
	• Collected ads based on only some bad keywords
	• Analyses historical scam data and its relationship with economic, structural, and cultural characteristics of the communities that are exposed to fraudulent advertising.
	• Automobile: Cars + Trucks section
	  ○ By Dealer (out scope)
	  ○ By Owner
	  ○ Used Crawler for data collection for (**2,424,092**)
	  ○ **3.5** months and they divided
	  ▪ Flagged (**42,185**)
	  ▪ Unflagged
	  ○ **Flagged reasons:**
	  ▪ CB=Census Bureau;
	  ▪ FBI= FBI Crime Database
	  ○ **They found that**
	  ▪ Not all scams are flagged
	  ▪ Not all flagged are scams
		• The study is limited to 30 American cities.
		• Thus, the results are likely not generalizable.
		• Individuals that buy automobiles online may not be representative of other fraudulent transactions enabled by Craigslist scams

Table 2.1 (continued)

Youngsam Park, Jackie Jones, Damon McCoy, Elaine Shi, and Markus Jakobsson, Scambaiter(2014), Understanding Targeted Nigerian Scams on Craigslist", NDSS '14, 23-26, San Diego, CA, USA, ISBN 1-891562-35-5	• Focus on fake payment scams targeting users on Craigslist
	  ○ Larger Organizations.
	  ○ Locations of Scammers.
	  ○ Methods and Tools.
	  ○ Email Account Usage.
	  ○ Filtering Messages.
	• built an automated data collection system
	  ○ Relied magnetic honeypot advertisements (Attractive)
	  ○ Using that, they offered goods for sale on Craigslist
	  ○ Gather three months of data and perform an in-depth analysis on;
	• Understanding of scammers' action patterns
	  ▪ **Automation tools:** Could not know much about them
	  ▪ **Scammers' email account usage**
	  • **Geolocation**.
	  ▪ Found 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.
		• ***Larger Organizations:*** 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.
		• ***Locations of Scammers:*** These groups use shipping addresses in both Nigeria and the U.S, but they did not validate the existence of those addresses.
		• ***Methods and Tools:*** The tools are fairly limited and include static text in the body of the reply or cannot parse subjects of listings.
		• ***Email Account Usage:*** Set the reply-to address in the initial message to a different email address that is reused often and longer-lived.
		• ***Filtering Messages:*** Lack of filters that are more effective at detecting scam messages.

Table 2.1 (continued)

Youngsam Park, Damon McCoy and Elaine Shi "Understanding Craigslist Rental Scams", Proceedings of Financial Cryptography and Data Security Conference (FC), Barbados, February 2016.	• Conduct the The first systematic empirical study of the online rental scams of the Craigslist rental section. • Our crawler revisited each crawled ad three days after the first visit to detect if they have been flagged by Craigslist. • Payment Request Analysis: Western Union or MoneyGram • Daily crawls of rental sections on Craigslist across 20 different cities and areas in the United States • Collected over 2 million in 141 days and 6.1% were flagged	• Unable to provide an estimate of how many scam listings they did not detect. • Investigate automated detection approaches to improve filtering (flagging rate) • Plan to perform test purchases from merchants to understand which banks they are contracting with to process credit card transactions • Many of these scams found were specific to the United States, while Craigslist is used in other countries.

Table 2.1 (continued)

Alan McCormick and William Eberle, Discovering Fraud in Online Classified Ads, May 2013	• Presented an approach for discovering fraud in classified ads. • Focus on applying data mining techniques to discover patterns and relationships in classified ad data. • Examine data provided by a classified ads website.	• Plan to modify our data mining approach to extract other relevant features from the advertisement data. • Create a tool to detect fraudulent ads the instant they are placed.

Summary

This chapter has detailed in online scams in general and craigslist notably. Scammers act both buyers and sellers, and that makes this issue very complicated. This chapter has covered the background related to the problem and solution, making our research's contributions very clear.

Chapter III: Methodology

Introduction

This chapter outlines and explains the research methodology to be used during this report, including the study's design, data collection, tools, and techniques. In this chapter, followed by the theory, the research approach was discussed to explain the success of certain activities required for data collection.

Besides, the discussions given in this chapter are focused on the theoretical context discussed in Chapter 1 and guided by Chapter 2, which includes reviewing the literature's analysis. The subject and study questions are the key drivers governing the researcher to settle on a suitable methodology for collecting data. Thus, the following sections will briefly address the steps involved in the design of operational research to achieve the intended objectives and the anticipated outcome of the report.

Design of the Study

We'll be taking the following steps in this study;

1. Identify as much as possible of the scam triggers list. It will involve studying the scam perspectives of Craigslist buyers and sellers. In this stage, the experience of the customer will be the key driver for the recognition of scam triggers.
2. Find all Craigslist related ads related to the triggers found in the first phase. In this step, all the adds that suit our list of scam triggers will be collected. One ad found, then it will be held for a couple of days to see what happens according to the Craigslist rules for those ads.

3. Analyze advertisements based on how many have been detected or deleted, and this would have been deemed to be fraud the most. All ads may be either flagged, added, deleted, or stayed the way it is. All we care about in this phase is how many ads were flagged, hidden, or deleted, and then we find them to be a scam.

Data Collection

This report aims to gather and check as many triggers as possible to alert the sellers and buyers of Craigslist not to be scammed. We may recognize several cases where the probability of it being a scam is high for someone who might use these in their ads. These stimuli are going to be used to classify advertisements and evaluate the possibility of an ad that it's a scam. This explains how scammers act and interact over the Craigslist, which can dramatically reduce the risk of scamming users (buyers and sellers). We'll make it safer and more confident, though; it's to ensure that those stimuli work.

To make the job easier for them (Buyers & Sellers), we're going to recommend ways to automate the suspect keyword searches. We will need to use a scraping tool or manual search to crawl a Craigslist website like scam-detector.com for the bad keywords to find the keywords.

We will describe the list scam triggers at this point, and collect our data from Craigslist based on those triggers only. Instead, whether the causes are actual scam or not using the flag method available in Craigslist, they will be checked. Finally, we'll introduce how many of those scam causes are likely to be actual fraud.

Tools and Techniques

One of the anonymous online U.S. advertising produced in 1995 is the Craigslist website (https:/www.craigslits.org), which is almost free for any human. Craigslist is available worldwide from all 50 U.S. states and nearly 500 cities. Approximately 60 million users have been recorded using Craigslist and have around 50 billion views of the website each month, with almost 13 different languages. It is also worth noting that there are nearly 80 million commercials per month. With all the above-mentioned benefits, there are certainly practically unlimited bad guys around to be of use to it and abuse the whole services to scam buyers and sellers alike.

This work would be the first of its kind to provide a range of scam triggers (identifiers) for some of those scammer's tactics and use those triggers against Craigslist ads to verify certain ads whether they are scam or not and that will be a great benefit for both buyers and sellers. This service is just people buying and selling personal belongings without any real mediator between them. This makes finding the bad guys very easily very difficult. It is also very convenient for the bad guys to use this service to trick people.

They should all recognize that Craigslist aims to minimize these scams as much as possible, but with the high number of users and advertisements, removing these scammers is very difficult. Flagging program is one of Craigslist's best services available to its users. It allows every user the option to report every questionable ad that Craigslist's team will review quickly. The owners will negotiate with the Craigslist team with any flagged ads and have more evidence that their flagged ads are not a fraud. If,

by some chance, the owner of such flagged ads refuses to show that the ad is not fraud, then that ad will be immediately removed (deleted) from Craigslist entirely.

It is also worth noting here that there can be a misuse of the flagging system for many reasons and one of them if someone competes with someone, and they can flag ads from those with whom they compete. Craigslist team will reach out to those ad owners to ensure they're not scamming until their ads are filtered or removed. So not all flagged ads in this study should be considered fraud, but all disabled (deleted) ads are considered fraud.

Summary

This chapter deals in detail with what the analysis approach covers. It has been discussed the design, which will be in three stages of discovering the scam triggers, validating those triggers against Craigslist ads, and sorting those ads based on flagged, then removed ads. Craigslist will be the primary data source explored for this study. Finally, the methods and procedures used to gather and validate data have also been clearly defined.

Chapter IV: Data Presentation and Analysis

Introduction

This chapter offers a comprehensive presentation and review of data that we manually collected from the craigslist. Craigslist ads are the only source of data that will be based or used in this study here to support our earlier defined scam triggers. The data collected or advertisements are restricted to section only automotive cars & tracks, then either proprietary or dealer sub-sections in craigslist. The data was also geographically limited to only four U.S. cities, which are: Georgia's Atlanta, Houston or Texas, California's Los Angles and Washington's Seattle.

Data Presentation

The research has identified seven scam triggers in which ads available in Craigslist will be detected to collect our data. Then the analysis will test those triggers to show how likely such triggers are to be a scam. The scam triggers the focus of the research comes directly from the experiences of people dealing with Craigslist advertising, whether they are buyers or sellers, and warned others to be wary of those advertising with those identifications. The list of scam triggers and a brief explanation of each are as follows;

- ***Bad Keywords Scam Trigger:*** In this trigger, we've collected all the ads with all the bad keywords we name in the description (details) of the ads such as; Army, Military, PayPal, Western Union, and Cashier's Check from all the four above cities.

- ***Personal Emails Scam Trigger:*** In this scam trigger, we have collected all the above four cities all ads that have a personal email in the definition of the ads (details), such as; Hotmail, Gmail, and Yahoo.
- ***Multiple Location Posting Scam Trigger:*** We also discovered some ads during our analysis posted multiple times within the same location or different locations. Still, it has a slide or minor differences either in the ad name or in the price.
- ***Rouge Picture Scam Trigger:*** In this scam trigger, we have collected all advertisements which have pictures taken from other sources in the advertising. The primary source from which it was copied is google.com.
- ***Voice Over IP Scam Trigger:*** In this scam trigger, we collected all ads in the advertising which have voice over IP phone numbers. Voice over IP phones can be set up online, and that phone owner is hard to track.
- ***Too Good to be True Scam Trigger:*** We tested each car price in this trigger and compared it with market value. It isn't going to take too much time today to know the market value for a specific vehicle.
- ***Dealers Posting as Owner Scam Trigger:*** We have found some ads in this trigger posted as an owner but posted by a dealer and then added those ads to our watchlist. We have seen this as they have inserted a phone number, which is the dealers' phone, or they could have two posts with the dealer in one and the owner posting in the other.

Data Analysis

It was collected in our data based on matching one or more of the scam as mentioned above triggers, which formed the basis for our data collection. -- ad to be presented here has at least one scam trigger, which was the minimum requirement to collect our data or ads in Craigslist form. If. The ad has one or more scam trigger character; we've put the following details in our datasheet;

- ***Ad Location:*** This means where the commercial was released. Posters can post their ad on Craigslist, depending on where they need to sell it. Our data collection has confined our data in only four major cities in United Stated: Georgia's Atlanta, Texas' Houston, California's Los Angeles, and Washington's Seattle.
- ***Scam Trigger Type:*** As discussed above, we have seven scam triggers that will be used to get or collect the ads. In several forms of scam triggers, each ad may have at least one or more in either the Ads title or the ad description.
- ***Ad Title:*** Each ad has a short title that often reflects a brief view of the ad, which is very useful to both sellers and buyers. That is what mostly attracts buyers to have an interest in the advertisement from the quick view of the ad title, and we find that to have a concise and appealing ad title is, and should be, crucial or essential for all buyers.
- ***Ad URL:*** We also kept each ad URL in our records, because in a few days, we wanted to revisit the same ad.
- ***Ad Status:*** This comes just a few days after the ad has been written. We care more about this than anything else because if the ad is likely to be a scam, then some

people may quickly realize that the ad is not normal, and then they may flag the ad. Then Craigslist's team will be closely watched to ensure that the ad is normal, or they may delete the ad if they don't just get what they need from the ad poster or owner. This is not under our control; we track the changes and then record and present them. Our data has, so far, recorded five different types that are;

- ***No status change:*** Once the ads were observed, then it stayed the same with no change happened at all.
- ***Post not found:*** This could happen two ways; either post has been removed quickly from Craigslist team, or the ad was sold out and is no longer available.
- ***Deleted by the author:*** This seems clearly something went wrong with the ad, and the owner has been removed the ad.
- ***Post expired:*** This is because the post has reached its display threshold days, then the add won’t be displayed and needs to be reposted again.
- ***Flagged for removal:*** This means something is wrong with the post or the ad, and presumably, someone has reported proactively and so the ad owner might have reached out to Craigslist team and then put the ad ready to be deleted. This could stay for a few more days until it is permanently deleted and the post will not be visible anymore.

The following data summary table presents the entire data summary which means each scam trigger and its status for quick view;

Table 5.1

Data Analysis Summary

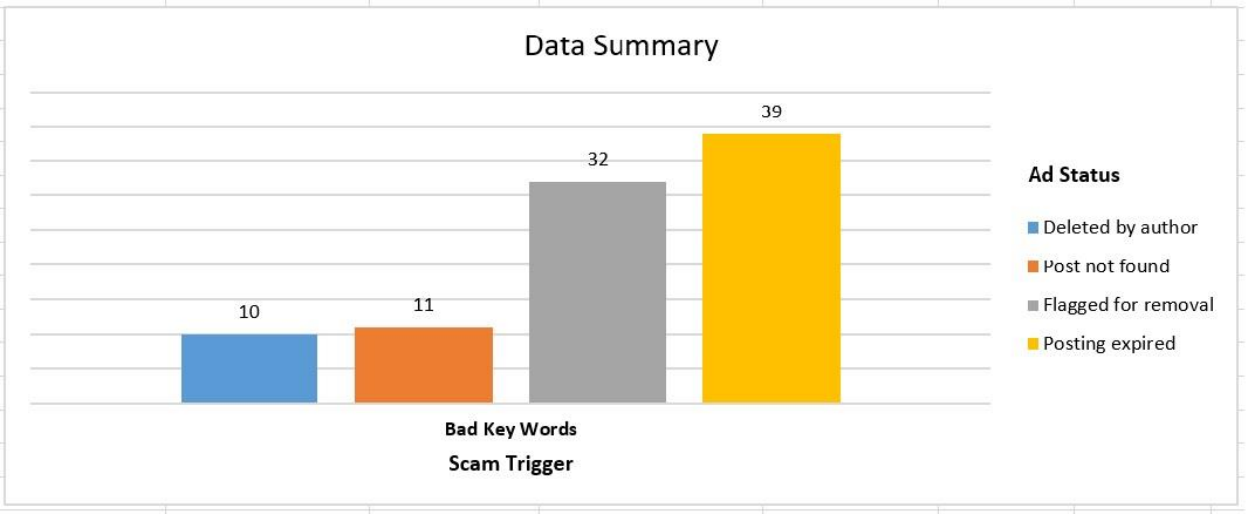
Scam Trigger	**Ad Status**					
	No Status Change	Post not found	Deleted by author	Posting expired	Flagged for removal	**Total**
Bad Key Words		11	10	39	32	**92**
Dealers Posting as Owner	2	2	1	6	25	**36**
Multiple Location Posting					26	**26**
Personal Email	16	5	28	13	71	**133**
Rogue Picture			18			**18**
Too Good to be True	1	2	1	6	25	**35**
Voice Over IP	1	2	1	4	14	**22**
Total	20	22	59	68	193	**362**

With the above data summary in a table format, the following is the representation of the data in a graph, and that will consist of the Scam Trigger, Number of Ads, and the Status of each Ad in a chart.

Bad Keywords Scam Trigger: According to our report, we have 92 ads that have bad keywords in it, which constitutes 25.4 percent of all of our advertising. In a closer look at that data, we found that 11 ads were not found in Post status, that its author had removed ten ads, that 39 ads had expired and that 32 ads had been marked for removal. Graphical representation of the data is as follows;

Figure 5.1

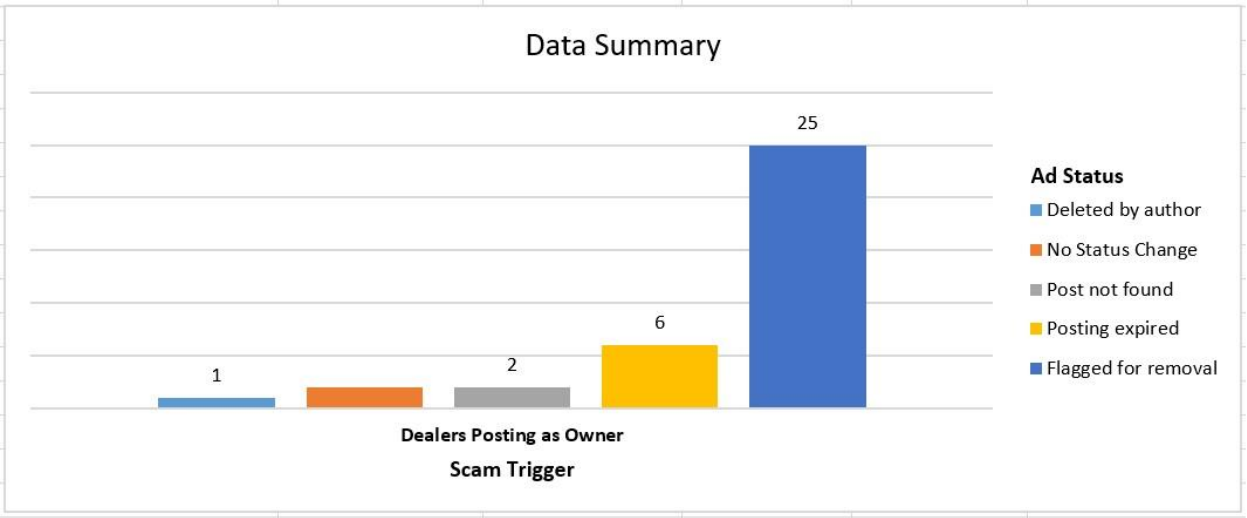
Bad Key Words Scam Trigger



Dealers Posting as Owner Scam Trigger: According to our data, we have 36 ads posted as owner by dealers, representing 9.9% of our data. In a closer look at that data, we found that two ads were not found in Post status, while two other ads had no changes in post status at all, its owner removed one ad, six ads had expired, and 25 ads had been marked for removal. Graphical representation of the data is as follows;

Figure 5.2

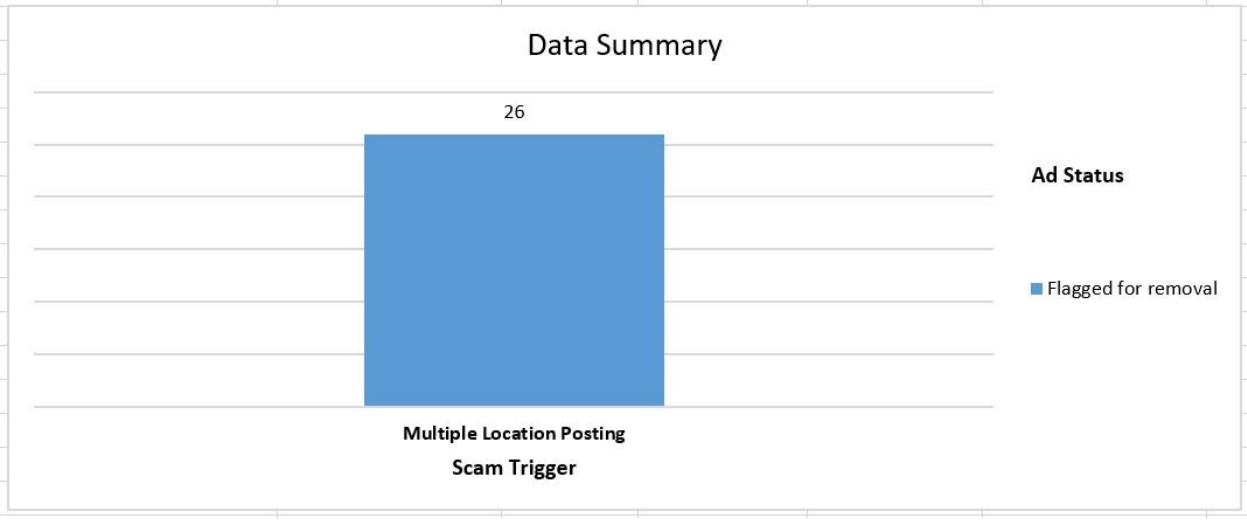
Dealers Posting as Owner Scam Trigger



Multiple Location Posting Scam Trigger: According to our data, we have 26 ads posted several times at the same location in a different location, and this data represents 28.2% of all our data. After observing the results, we discovered that the entire 26 ads were flagged for removal in a very short period. Graphical representation of the data is as follows;

Figure 5.3

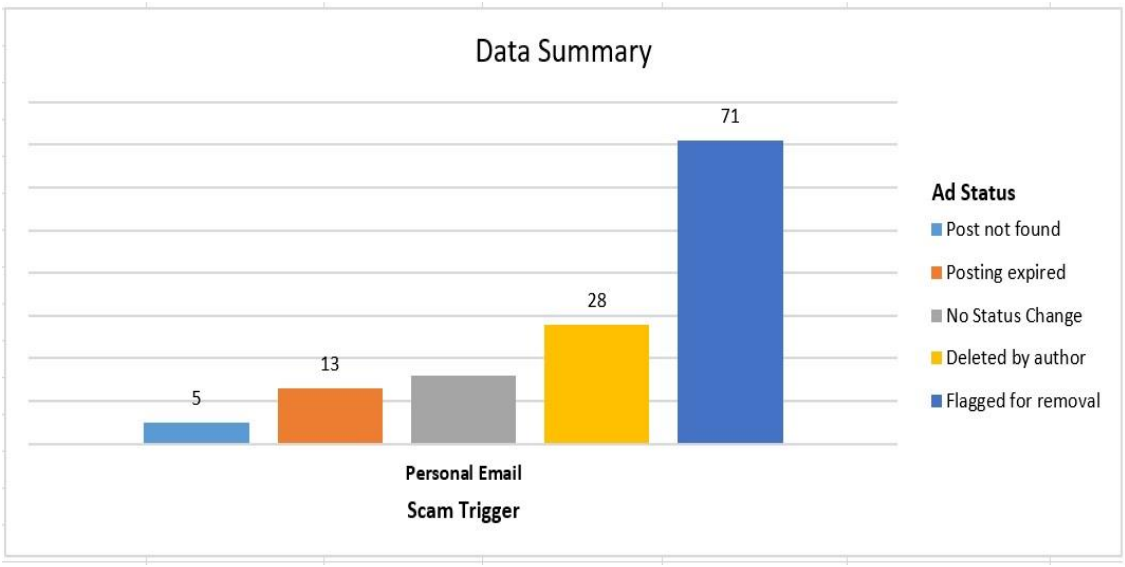
Multiple Location Posting Scam Trigger



Personal Email Scam Trigger: This scam trigger has the largest ads compared to the rest of the scam triggers, comprising 36.7 percent of all our results. In a closer look at that data, we found five ads in Post not found status, their author removed 28 ads, 13 ads expired while 16 ads did not alter the post status, and 28 ads were marked for removal. Graphical representation of the data is as follows;

Figure 5.4

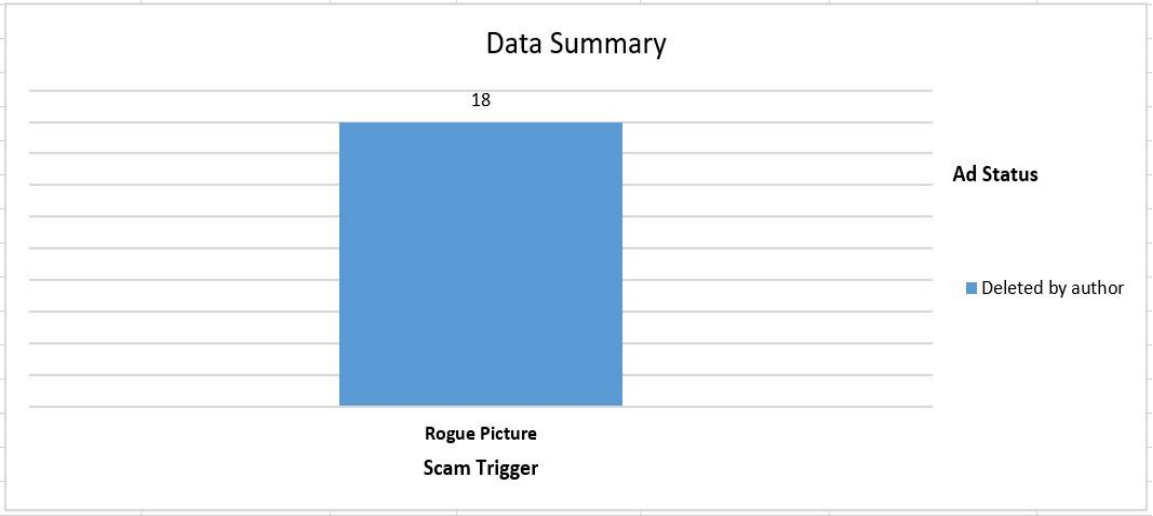
Persona Email Scam Trigger



Rogue Picture Scam Trigger: According to our data, we have 18 ads featuring rogue images, comprising 4.9 percent of our results. After revisiting the ads a few days later, we found that its author deleted all 18 ads means the ad owner. Graphical representation of the data is as follows;

Figure 5.5

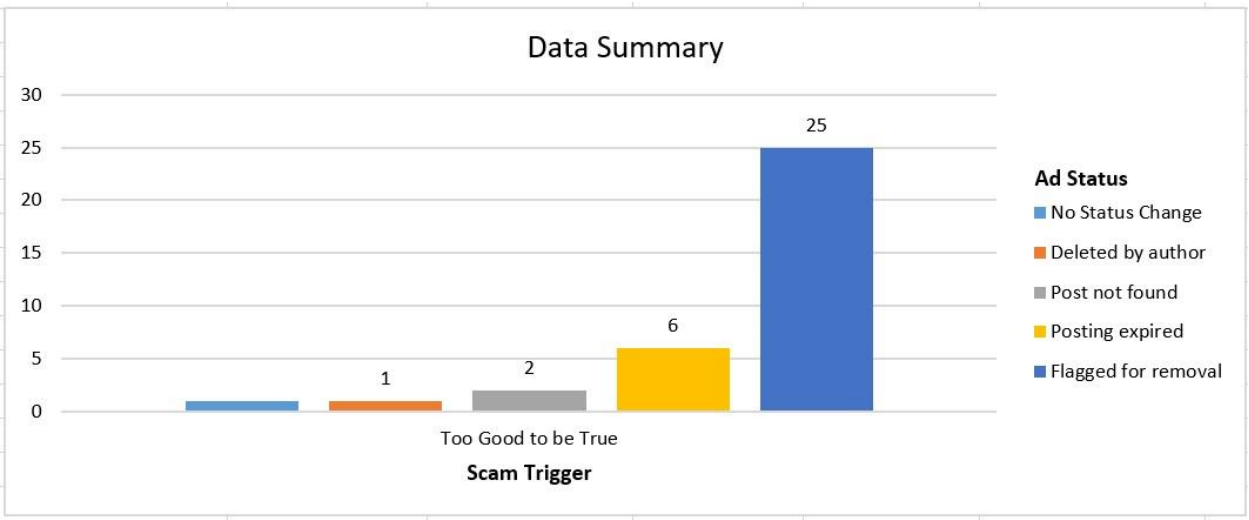
Rogue Picture Scam Trigger



Too Good to be True Scam Trigger: According to our data, this ad has 35 ads that have too good to be true scam trigger in it, and that represents 9.6 percent of all our data. After a few days, we found two ads in Post not found status, its owner removed one ad, one ad did not change state, six ads expired, and 25 ads were marked for removal. Graphical representation of the data is as follows;

Figure 5.6

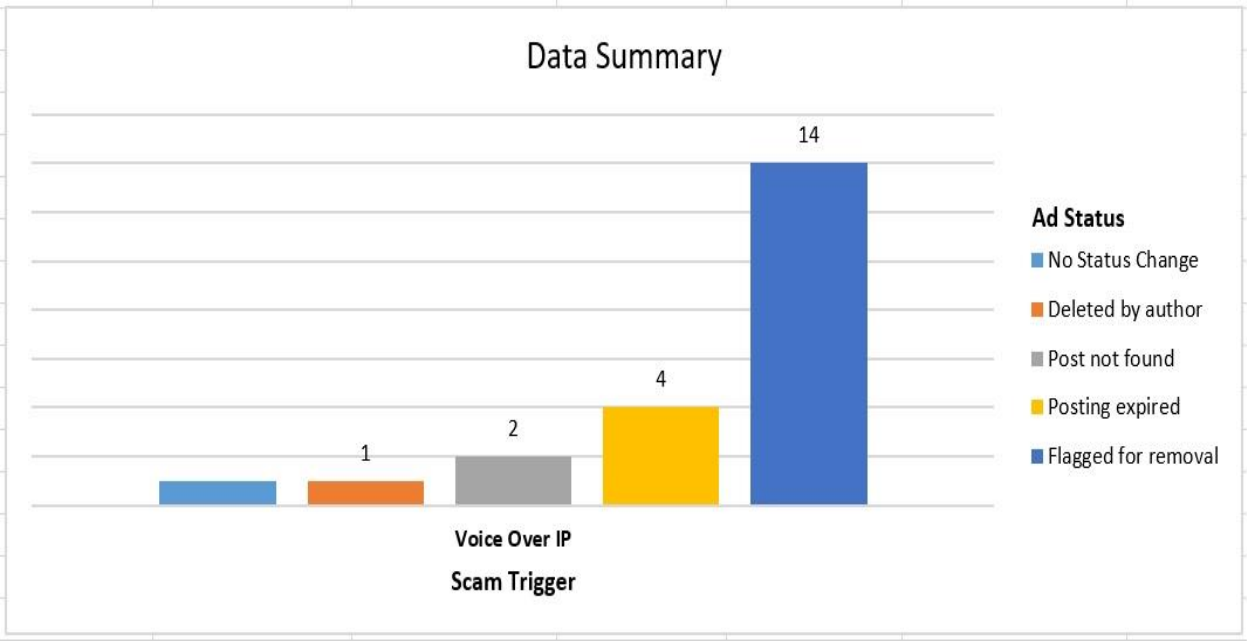
Too Good to be True Scam Trigger



Voice Over IP Scam Trigger According to our results, this ad has 22 ads that have voice over IP scam trigger in it, and that represents 6.0 percent of our total data. After a few days, we found two ads in Post not found status, its owner removed one ad, one ad did not change state, four ads expired, and 14 ads were marked for removal. Graphical representation of the data is as follows;

Figure 5.7

Voice Over IP Scam Trigger



Summary

This chapter mainly and deeply covered all of our results. We included a full description of our data and its part in the data presentation, which covered how we gathered all of our data with a scam trigger and then collected ads from craigslist based on that scam trigger. Every ad collected has one or more scam triggers that we had identified before we looked at those ads and found out. In the section on data analysis, we discussed a detailed research result. The result shows all the advertisements we received, then what happened after observing a few days of each ad and then

introducing each scam trigger, number of ads received for that scam trigger, and what occurred per ad.

Chapter V: Results, Conclusion, and Recommendations

Introduction

The chapter will explore the results, conclusions, and future works. The findings presented in this chapter should concentrate on the research goals and the problem set out in chapter one. These findings should be a direct response to the research question, as mentioned earlier.

Results

Research questions that were identified in chapter 1 will be answered here in this result discussion. Our objective and answers will be as following;

Question 1: What are the scam triggers to detect online scams on craigslist?

To answer the above question, and as the research discussed in depth in our earlier chapter, we have identified seven scam different scam triggers as following;

- Bad keywords scam trigger.
- Dealers posting as owner scam trigger.
- Multiple location posting scam trigger.
- Personal email scam trigger.
- Rogue picture scam trigger.
- Too good to be true scam trigger.
- Voice over IP scam trigger.

The research has identified the above seven scam triggers and discussed them in detail in an earlier chapter. The following will be our next object;

Question 2: How to find and analyze ads based on identified scam triggers?

To answer the above question, the research has able to identify 362 ads from craigslist based on our scam triggers. The research, as discussed earlier, categorized ads based on scam triggers, and each ad was given a status, which was what happened to the ad couple of days after.

As discussed in an earlier chapter, the following is the last objective of this study;

Question 3: What the likelihood to be considered those ads as a scam?

The research found 362 ads, and we have given the following statuses, and that will show the likelihood to be considered those ads as a scam. Keep in mind those statuses were given after revisiting those ads a couple of days later, and the following are those statuses;

- **No Status Change:** According to our result, we found out 5.5% of our ads have no status changes at all, which means those ads remained as they were posed.
- **Post not found:** Our research shows that 6.0% of our ads have been removed very quickly or ad simply disappeared.
- **Deleted by the author:** The results show that 16.2% of our ads have been deleted by the author or the owner of the ad. That could happen either. They simply deleted or reposted again and deleted that particular posting.

- **Posting expired:** Our data show that 18.7% of those ads have simply expired, which means it has reached the maximum posting days required, and the owner may need to repost the ad again.
- **Flagged for removal:** Our data shows that 53.3% of the ads have been flagged for removal. This is a clear indication that some users have proactively flagged those, and then it has been placed those ads to be removed. Those ads may stay as flagged for removal status before they were permanently removed.

Conclusion

The research was able to successfully answer all the research questions in this report. This research's main objective was to discover or list scam triggers or indicators and then figure out whether there will be any ads that have those scam triggers from craigslist characteristics and eventually prove the probability of those ads to be classified as spam.

Future Work

While the research has been effective in identifying seven specific scam triggers or indicators and has also been able to identify ads from craigslist based on those scam triggers and has given the probability of scam trigger being scam based on the number ads and what happened to those ads a few days later, the research looks at a simple future work on the following;

- Scam triggers can be expanded and added more into that list as time permits.

- All identified ads can be studied more on their characteristics and build an automated grading system that gives the likelihood of those ads to considered as a scam.

References

Abernethy, J., Chapelle, O., & Castillo, C. (2008). Web spam identification through content and hyperlinks. *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web - AIRWeb '08*, 41. https://doi.org/10.1145/1451983.1451994

Alan McCormick and William Eberle, Discovering Fraud in Online Classified Ads, May 2013

Alexa. 2012. http://www.alexa.com.

Anderson, R., Barton, C., Bohme, R., Clayton, R., Ganan, C., Grasso, T., Levi, M., Moore, T., & Vasek, M. (2019). *Measuring the Changing Cost of Cybercrime*. 32.

Aransiola, J. O., & Asindemade, S. O. (2011). Understanding cybercrime perpetrators and the strategies they employ in Nigeria. *Cyberpsychology, Behavior and Social Networking*, *14*(12), 759–763. https://doi.org/10.1089/cyber.2010.0307

Aston, M., McCombie, S., Reardon, B., & Watters, P. (2009). A Preliminary Profiling of Internet Money Mules: An Australian Perspective. *2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*, 482–487. https://doi.org/10.1109/UIC-ATC.2009.63

Benczúr, A. A., Bíró, I., Csalogány, K., & Uher, M. (2006). Detecting nepotistic links by language model disagreement. *Proceedings of the 15th International Conference on World Wide Web - WWW '06*, 939. https://doi.org/10.1145/1135777.1135954

Bíró, I., Szabó, J., & Benczúr, A. A. (2008). Latent dirichlet allocation in web spam filtering. *Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web - AIRWeb '08*, 29. https://doi.org/10.1145/1451983.1451991

Buchanan, J., & Grant, A. J. (2001). Investigating and Prosecuting Nigerian Fraud. *United States Attorneys' Bulletin*, *49*, 39.

Buchanan, T., & Whitty, M. T. (2014). The online dating romance scam: Causes and consequences of victimhood. *Psychology, Crime & Law*, *20*(3), 261–283. https://doi.org/10.1080/1068316X.2013.772180

Clayton, R., Moore, T., & Christin, N. (2017). *Concentrating Correctly on Cybercrime Concentration*. 34.

CraigsList. 2012. Craigslist Factsheet. http://www.craigslist.org/about/factsheet

Fetterly, D., Manasse, M., Najork, M., & Avenida, L. (2004). *Spam, Damn Spam, and Statistics: Using statistical analysis to locate spam web pages*. https://doi.org/10.1145/1017074.1017077

Florencio, D., & Herley, C. (2015). *Microsoft Research, One Microsoft Way, Redmond, WA, USA dinei@microsoft.com, cormac@microsoft.com*. 5.

Garg, V., & Niliadeh, S. (2013a). Craigslist Scams and Community Composition: Investigating Online Fraud Victimization. *2013 IEEE Security and Privacy Workshops*, 123–126. https://doi.org/10.1109/SPW.2013.21

Garg, V., & Niliadeh, S. (2013b). Craigslist Scams and Community Composition: Investigating Online Fraud Victimization. *2013 IEEE Security and Privacy Workshops*, 123–126. https://doi.org/10.1109/SPW.2013.21

Herley, C. (2015). *Why do Nigerian Scammers Say They are from Nigeria?* 14.

Huang, J., Stringhini, G., & Yong, P. (2015). Quit Playing Games with My Heart: Understanding Online Dating Scams. In M. Almgren, V. Gulisano, & F. Maggi (Eds.), *Detection of Intrusions and Malware, and Vulnerability Assessment* (pp. 216–236). Springer International Publishing. https://doi.org/10.1007/978-3-319-20550-2_12

Imposter Scams Top Complaints Made to FTC in 2018. (2019, February 28). Federal Trade Commission. https://www.ftc.gov/news-events/press-releases/2019/02/imposter-scams-top-complaints-made-ftc-2018

Infographic: Internet Scamming is on The Rise. (n.d.). Statista Infographics. Retrieved November 7, 2019, from https://www.statista.com/chart/15069/number-of-internet-scams-in-the-us/

Isacenkova, J., Thonnard, O., Costin, A., Balzarotti, D., & Francillon, A. (2013). Inside the SCAM Jungle: A Closer Look at 419 Scam Email Operations. *2013 IEEE Security and Privacy Workshops*, 143–150. https://doi.org/10.1109/SPW.2013.15

Jay M. Ponte and W. Bruce Croft (2019), "A Language Modeling Approach to Infromation Retrieval" Retrieved November 8, 2019, from http://ciir.cs.umass.edu/pubfiles/ir-120.pdf

Johnson, C. (2014). Fakers, Breachers, Slackers, And Deceivers: Opportunistic Actors During the Foreclosure Crisis Deserve Criminal Sanctions. *Capital University Law Review*, 46.

Jones, J., & McCoy, D. (2014). The check is in the mail: Monetization of Craigslist buyer scams. *2014 APWG Symposium on Electronic Crime Research (ECrime)*, 25–35. https://doi.org/10.1109/ECRIME.2014.6963162

Kanich, C., Kreibich, C., Levchenko, K., Enright, B., Voelker, G. M., Paxson, V., & Savage, S. (2008). *Spamalytics: An Empirical Analysis of Spam Marketing Conversion*. 12.

Konte, M., Feamster, N., & Jung, J. (2009). Dynamics of Online Scam Hosting Infrastructure. In S. B. Moon, R. Teixeira, & S. Uhlig (Eds.), *Passive and Active Network Measurement* (Vol. 5448, pp. 219–228). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-00975-4_22

Kroft, K., & Pope, D. G. (2014). Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist. *Journal of Labor Economics*, *32*(2), 259–303.

Lengle, B., Sam, M., Lam, J., & Lee, A. (n.d.). *Security Analysis on Craigslist (December 2009)*. 5.

Levchenko, K., Pitsillidis, A., Chachra, N., Enright, B., Felegyhazi, M., Grier, C., Halvorson, T., Kanich, C., Kreibich, C., He Liu, McCoy, D., Weaver, N., Paxson, V., Voelker, G. M., & Savage, S. (2011). Click Trajectories: End-to-End Analysis

of the Spam Value Chain. *2011 IEEE Symposium on Security and Privacy*, 431–446. https://doi.org/10.1109/SP.2011.24

Martinez-Romo, J., & Araujo, L. (2009). Web spam identification through language model analysis. *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web - AIRWeb '09*, 21. https://doi.org/10.1145/1531914.1531920

MaxMind. 2012. Fraud Detection through IP Address Reputation and a Mutual Collaboration Network. (http://www.maxmind.com/Maxmind_WhitePaper.pdf)

McCormick, A., & Eberle, W. (2013). *Discovering Fraud in Online Classified Ads*. 6.

Moore, T., Clayton, R., & Anderson, R. (2009). The Economics of Online Crime. *Journal of Economic Perspectives*, *23*(3), 3–20. https://doi.org/10.1257/jep.23.3.3

National Consumers League, 2012, http://www.nclnet.org/.

Nilizadeh, S., Orlova, D. A., Nematzadeh, A., Gupta, M., & Kapadia, A. C. (2011). *In Search of a Toyota Sienna: Prevalence and Identification of Auto Scam on Craigslist*. 22.

Ntoulas, A., Najork, M., Manasse, M., & Fetterly, D. (2006). Detecting spam web pages through content analysis. *Proceedings of the 15th International Conference on World Wide Web - WWW '06*, 83. https://doi.org/10.1145/1135777.1135794

Ojedokun, U. A., & Eraye, M. C. (2012). *Socioeconomic Lifestyles of the Yahoo-Boys: A Study of Perceptions of University Students in Nigeria*. *6*(2), 13.

Park, Y., Jones, J., McCoy, D., Shi, E., & Jakobsson, M. (2014). Scambaiter: Understanding Targeted Nigerian Scams on Craigslist. *Proceedings 2014 Network and Distributed System Security Symposium*. Network and Distributed System Security Symposium, San Diego, CA. https://doi.org/10.14722/ndss.2014.23284

Park, Y., McCoy, D., & Shi, E. (2017). Understanding Craigslist Rental Scams. In J. Grossklags & B. Preneel (Eds.), *Financial Cryptography and Data Security* (Vol. 9603, pp. 3–21). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-54970-4_1

Piskorski, J., Sydow, M., & Weiss, D. (n.d.). *Exploring Linguistic Features for Web Spam Detection: A Preliminary Study*. 7.

Policing Cybercrimes: Situating the Public Police in Networks of Security within Cyberspace. (2010). 36.

PriceWaterhouseCoopers. 2012. Internet Advertising Revenue Report. sponsored by The Internet Advertising Bureau, 2011 Full Year Results, April 2012.

Rege, A. (2009). *What's Love Got to Do with It? Exploring Online Dating Scams and Identity Fraud*. *3*(2), 19.

Smith, A. (2009). Nigerian Scam E-Mails and the Charms of Capital. *Cultural Studies*, *23*(1), 27–47. https://doi.org/10.1080/09502380802016162

Tran, H., Hornbeck, T., Ha-Thuc, V., Cremer, J., & Srinivasan, P. (2011). Spam detection in online classified advertisements. *Proceedings of the 2011 Joint*

WICOW/AIRWeb Workshop on Web Quality - WebQuality '11, 35. https://doi.org/10.1145/1964114.1964122

Urvoy, T., Chauveau, E., Filoche, P., & Lavergne, T. (2008). Tracking Web spam with HTML style similarities. *ACM Transactions on the Web*, *2*(1), 1–28. https://doi.org/10.1145/1326561.1326564

Vaibhav Garg and Shirin Nilizadeh, "Craigslist Scams and Community Composition: Investigating Online Fraud Victimization", 2013 IEEE Security and Privacy Workshops, DOI 10.1109/SPW.2013.21

Youngsam Park, Damon McCoy and Elaine Shi "Understanding Craigslist Rental Scams", Proceedings of Financial Cryptography and Data Security Conference (FC), Barbados, February 2016.

Youngsam Park, Jackie Jones, Damon McCoy, Elaine Shi, and Markus Jakobsson, Scambaiter (2014), "Understanding Targeted Nigerian Scams on Craigslist", NDSS '14, 23-26, San Diego, CA, USA, ISBN 1-891562-35-5

Zollman, P. 2012. Craigslist 2012 revenues increase 9.7%. AIM Group. http://aimgroup.com/2012/11/07/craigslist-2012-revenues-increase-9-7-big-four-battle-for-global-classified-lead/