Old is Still Gold: A Comparison of Cyber and Regular Consumer Fraud in The United States

Mohammad Taha Khan and Chris Kanich
Department of Computer Science, University of Illinois at Chicago
taha@cs.uic.edu, ckanich@uic.edu

***Abstract*—Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.**

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

I. Introduction

In the United States, a total of more that 25 million people are victims of frauds [**?**]. These deceptive scams are a major cause of users economic harm with the added externalities of wasted time and stress. These illicit practices have a part of the underground economy for a while. Initially individuals were tricked by scam calls, mail or in-person fraudsters, however, the rise of the Internet has provided fraudulent entities with a more streamlined exposure to the overall population. A survey report from 2014 indicated that 47% of the Americans were victims of an online identity theft. Another recent report released by the Federal Trade Commission (FTC) revealed that debt collection, identity theft, and impostor scams contribute towards 56% of the total frauds complaints in 2015 [**?**] . With the number of Internet users on the rise, the number of cyber frauds is likely to increase over the next couple of years [**?**]. To deal with the increasing trend, along with the FTC, the Federal Bureau of Investigation (FBI) has also established a similar complaint portal known as the IC3, for the collection of specifically Internet-based fraudulent practices [**?**]. This emerging trend of deceptive practices necessitates their study in order to evaluate and mitigate the harm caused to victims individuals.

In this paper, we evaluate the nature of consumer fraud in the United States. Our work provides a comparison between cyber and well as regular frauds. We categorize cyber frauds as all those deceptive practices that victimize users online, while regular ones comprise of frauds that target individuals over the phone, with mail, or in-person. While we understand that cyber frauds are a major focus of today's research, our comparison based approach allows us to better understand how they differ from conventional frauds. It also enables us to independently evaluate trends in regular frauds and to see wether fraudsters are adopting online mechanisms to target more individuals. This combined analysis aids us to devise strategic suggestions to develop better fraud reduction methodologies.

To evaluate fraud trends, we use a primary dataset from of FTC complaints from the year 2013 and 2014. We also collect demographic information from the US Census Bureau [**?**]. In addition to data collection, we devise a calibration methodology to identify and separate in cyber frauds from the regular ones in the complaint dataset.

Our work provides three main contributions. First, we evaluate the distinctive trends prevalent in cyber and regular frauds in the dataset. This encompasses their reporting numbers and methods. The nature of frauds which are more common in each specific category and the insights on the fraudsters who carry out these specific deceptive activities. Secondly, we look at ethnic, age, education and employment demographics in each specific category and evaluate if certain individuals are more likely to report crimes. Finally, based on our findings we provide suggestive measures that can be taken into account by regulatory agencies to reduce the overall fraud in the United States.

The rest of the paper is structured as follows, section II provides a comprehensive overview of the relevant work. In section III, we elaborate features of the datasets used in our analysis along with a description of our calibration methodology. Section IV summarizes our findings from the data followed by our suggestions to regulatory agencies in V. We conclude our work in section VI and discuss avenues of future research.

II. Related Work

As our work evaluates both cyber as well as regular frauds, we provide related work that encompasses both of these categories. However, we elaborate more on recent research which focuses on cyber frauds as more individuals are victims of these crimes [**?**] due to the increased Internet usage trends for sensitive activities. Before the Internet became a primary

Data Field	Field Description
Agency Name	The complaint collection agencies associated with the FTC.
Zip code Information	The zip code of the victim and the fraudulent entity.
Contact Method	The primary channel used by the fraudulet entity to contact the victim e.g. Internet, phone, mail.
Fraud Description	A description of nature the fraud, and its type e.g. credit card, fake product, debt collection.
Fraud & Reporting Date	The dates when the fraud initially occurred and the date on which is was reported.

TABLE I
THE DESCRIPTION OF THE DATA FIELDS THAT WERE PRIMARILY USED IN THE FOR DATA CALIBRATION AND ANALYSIS

hub of economic and social activity, researchers measured [**?**] and developed techniques based on statistical models [**?**], [**?**], [**?**] to detect phone and credit card based frauds. In the past few years, research evaluations have shifted focus towards cyber activity [**?**], [**?**], [**?**], [**?**] due to its high rate and the increased potential for harm.

Even though, term *"cyber fraud"* is usually associated with Computer Science, its recent socio-economic impact has motivated researchers in Economics, Law, and Finance to explore solutions by incorporating methodologies specific to their areas. Ionescu et. al [**?**] characterize the types and sources cyber frauds in global digital networks, they link the increase of financial fraud to the prevalent usage of Internet services for financial management and transactions. The authors suggest the involvement of all stakeholders and employees through awareness and training for containing and mitigating fraud. Similarly, Howard et. al [**?**] study malicious code attacks against financial networks and suggest technical detection and mitigation techniques for financial infrastructure. [**?**] studies how the cyber criminals have several potential advantages over their opposing law enforcement agencies. They suggest some practical steps to even out the differential gap.

Due to an increase in the overall concern for online fraudulent activity, there has also been state-sponsored research that measures the impact of fraud. Smyth et. al [**?**] measure the extent of cyber fraud in Canada in 2011. Their work indicates that a major chunk of frauds does not get recorded and suggest a need for a sentinel record fraud data, similar to the FTC complaint center in the US.

Another significant area of research focuses on understanding the demographics of fraud victims. A recent FTC Report [**?**] uses complaint data to quantify complaint rates and across different ethnic and education groups in the US. [**?**] also look at how demographics effect the likelyhood of an individual to complain about fraud. Researchers in [**?**] provide a comprehensive survey report that sums the reactions of the victims of an online data breach. They categorize their results in different income, education, age, and ethnic groups. Such research aims to provide organizations with informed insight to better develop policies for consumer rights protection.

In comparison to previous research which individually look at either cyber of regular fraud, w work provides a unique angle of evaluation. We evaluate charectaristics for both types of cyber and regular frauds and their demographic trends.

III. DATA AND CALLIBRATION

In this section we explain the charectarisitcs of our datasets and the sources they were obtained from. We also provide insight into the essential data processing and callibration methodology that we incorporate to classify and filter the data for a fair evaluation of our questions.

A. Data Description

Description	Value
% Cyber Complaints	52.1
% Regular Complaints	47.9
Moth with Most Compliants	July 2013
Moth with Least Compliants	Feb 2013

TABLE II
T-TEST RESULTS FOR CYBER AND REGULAR FRAUDS IN DIFFERENT PERIODS

1) FTC Complaint Dataset: The primary dataset that we use for our evlauation is a corpus of the complaint logs filed at different collecitons agencies for the FTC during the months Jan, 13 to June, 14. The Dataset comprises of a total of 865K complaints aggregated for cyber as well as regular fraud instances during the time period. Table I shows the fields of the original dataset along with their description summary. For the purposes of brevity, we only include the fields that were used in our analysis.

2) US Census Datasets: Zip code information in our complaint dataset allows us to perform demographic analysis of the frauds. We obtain the demographic information associated with zip codes available at the US Census Bureau website [**?**]. The specific information that we collect is stated below:

- Population density per zip code
- Education and income data [1]
- Age statistics
- Race and ethinic information

As zip codes provide a low level granularity, to aggregate adjacent zip codes we obtain the Zip codes to the Metropolitinan Statistical Area (MSA) mappings from [**?**]. MSA are essentially groups of geographically connected zip codes that demonstrate strong social andd economic linkage. While there are more than 40,000 zip codes in the United States there are only 382 distinct MSAs [**?**].

B. Calibration Methodology

As the FTC dataset was aggregated for all fraud channels a major calibration step we perfrom is to tag each crime complaint as either cyber or regular. An associated challenge for this was our limited view of the fraud description. To

[1]We obtained education data was obtained from [**?**].

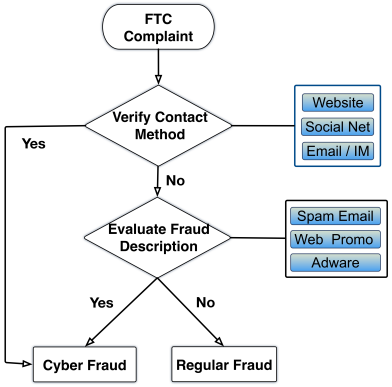

Fig. 1. Classification Methodology for Regular and Cyber Fraud

perform this calibration we use the **Contact Method** and **Fraud Description** fields from Table I. We flag a complaint as cyber if the victims primary contact method was thorugh online media, these are primarily websites, social networks, email and IM. For the remainder of the complaints we look at the description. If the complinat description invovles something associated with Internet, we classify it as a cyber fraud regardless of how the victim was initially approached. Figure 1 provides an depicticon of our classification methodology. This process provides us with two distinct categories and enables a fair comparison of the frauds.

IV. Evaluation

Our evaluation section can be divided into two main components. First, we look at the different aspects and trends within the complaint dataset and elaborate on certain prominent characteristics between regular and cyber frauds. The second part of our evaluation focuses on an in-depth analysis of the demographics linked with the fraud types.

A. Fraud Variation over Time

We initially perform a temporal analysis of the 15-month dataset to evaluate how the fraud reporting varies, over time and explore when a certain fraud is more likely to be reported. As a limitation of our dataset[2], we use reporting dates as an estimate of fraud count for a specific date. While the overall rate of fraud remains consistent, we observe significant variations in the winter holiday season. To investigate this, we select two distinct, 20 day periods in the dataset; we label them as **Working** (Aug, 15 to Sept, 5) and **Holiday** (Dec 15, to Jan, 5). Figure 2 shows the variation of cyber and regular frauds within the two specific time periods. We observe a significant drop in the frauds during the the holiday time period. Individually, cyber fraud decreases by 26% while regular fraud decreases by a much larger value of 56%. We belive that the larger decrease in regular frauds is an overesitmate as reuslt of a bias. We also compare cyber and regular frauds across each region by performing a the statistical t-test. *p-value*. Table III summarizes our results. We use the derived *p-value* coefficients aloing with additional insights from IV-B to support our intuition for the biased decrease in regular fraud reports.

[2]The dataset had singinifcant missing values for the date if inital contact of fraud, Figure 4 (a) shows that approximately 70% of reporitng dates were within a week of the date when the incident occoured.

Fraud Type	*p-value* Coefficient
Cyber	0.014
Regular	0.003

TABLE III
T-TEST RESULTS FOR CYBER AND REGULAR FRAUDS IN DIFFERENT PERIODS

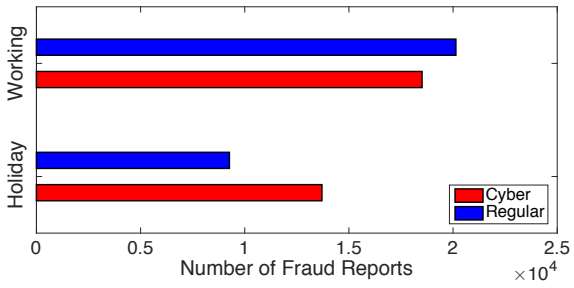

Fig. 2. Cyber and regular fraud variation during regular and holiday season

B. Fraud Reporting Methods

We evaluate the methods that individuals use to report fraud incidents to the FTC. We aggregate the 26 complaint collecting agencies into online and offline categories. For instance, reports made via the Internet complaint center or the FTC complaint assistant and tagged as online, while the ones made to the FTC call center, publisher clearing house, attorney generals or other regulatory institutions are categorized and offline. Figure 4 provides the distribution of how individuals opt to report cyber and regular crimes. Approximaltey 82% of cyber fraud victims used an online complaint facility, and 63% of regular fraud victims reported via offline methods.

A major chunk of regular frauds are reported to offline institutions which have reduced operation during holidays. We believe this significantly contributes to the reduction bias of regular frauds value in Figure 2. The statistically significant *p-value* for regular frauds in Table III, rejects our null hypothis of *identical reporting trends between two the two time periods*.

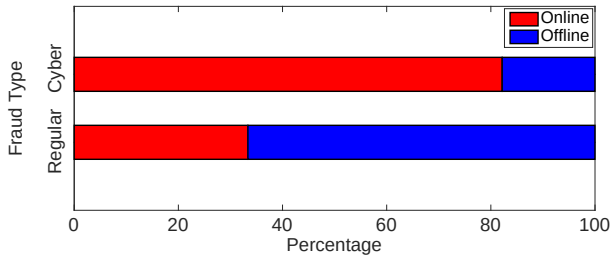

Fig. 3. Distribution of reporting methods in for cyber and regular frauds

To futher corroborate our reasoning, we evaluate an increase in reportings between the last 10 holidays and 10 working days right after. While cyber reports only incease by 37% the increase in regular fraud reports is a staggering 104%. The ratio of increase between each category is in line with the proportions of offline reporting methods used by victim individuals from each category. We belive that while both types of frauds experiecne a decrease, the decrease in cyber frauds represents a more accurate trend. These derived insights enable us to suggest more meaningful measures in section V to deal with consumer fraud.

C. Consumer Reaction Time

We evaluate how quick do the victims of a fraud report an incident. Figure 3 (a) shows the CDF of the number of days between the date when the fraud occoured and when in was first reported to the FTC. We do not observe any differene between the reaction time trends of cyber and regular fraud victims. The graph shows that for both fraud categories, almost 30% of the individuals take more than a week to respond. We beleive that this provides ample time window for fraudsters to maliciouly act on the assets aquired from individuals. This increased delay can also be a result of individuals discovering they were victimized to a fraud at a later date that the actual incedent. One example would be of a credit card theft, when the victims only realizes after they see an unauthorized transaction. Unfortunaltety, we do have enough information in the dataset to normalize againt this trend.

D. Most Common Frauds

Table IV provdies a summary of the top frauds in each category. The significant cyber presence of impostor scams, and sweepstackes in, which have long existed in the regular fraud domain provides evidence that fraudsters are adopting new technology to execute the same frauds. We incorporate this specicific insight in our discussion in section. Our results greatly match with the top sources of frauds in the FTC correlate with the frauds stated in an FTC news release in 2016 [**?**].

Cyber	%	Regular	%
Online Shopping & Sales	14.1	Impostor Fraud	29.8
Impostor Fraud	10.8	Telemarketing	20.1
Unsolicited Email	7.71	Debt Collection	16.1
Counterfeit Check Scams	7.40	Prizes & Sweepstakes	15.5
Prizes & Sweepstakes	7.40	Grants & Credit Loans	4.14

TABLE IV
TOP FRAUD TYPES

E. Top Fraudster Locations

Next, we identify the primary locations of the fraudsters within the United States and present our findings in reftopareas. While fraudulent entities are spread throughout, most of the heavy hitters belong to the metropolitan areas. We believe this provides an efficient disguise to the fraudulent entities. Numbers: 29.3% cyber reports and 32.6% regular reports.

Metropolitan Area (MSA)	% Cyber	% Regular
New York, New Jersey, Long Island	8.41	7.67
Los Angeles, Long Beach, Santa Ana	6.50	7.09
Washington, Arlington, Alexandria	4.10	5.78
Miami-Fort Lauderdale, Pompano Beach	4.16	5.40
Dallas, Fort Worth, Arlington	2.89	3.36
Chicago, Naperville Joliet	3.17	3.27

TABLE V
TOP METROPOLETAN AREAS WHERE FRAUTDSTERS ARE BASED

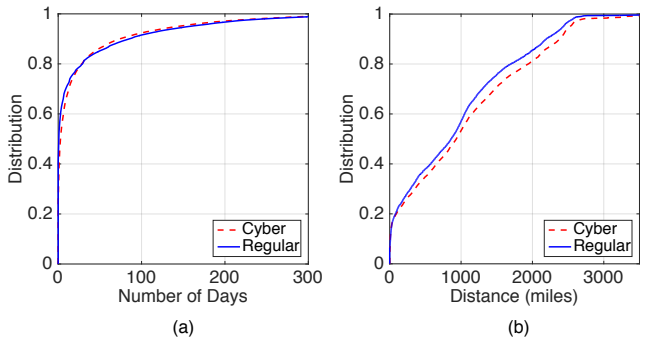

Fig. 4. Distribution of reporting methods in for cyber and regular frauds

We also observe certain areas that have a high cyber to regular fraud ratio and vice versa. The San Franciso, Oakland and San Jose, Santa Clara MSAs have a cyber to regular fraud ratio of 2.21 and 5.13. The popular fraud types in these regions are Internet services, unsolicited email, and online shopping. These areas serve as a good medium for cyber fraudsters as it allows them to gel into the surrounding cyber industry. Meanwhile, The Buffalo, Niagara Falls MSA has a regular to cyber fraud ratio of 8.76 with debt collection being the significant outlier. Further investigation reveals that Buffalo has a network of debt collecteors which have responsible for multi-million frauds [**?**], [**?**].

F. Fraudster Coverage

In order to understand the operational regions of fraudulent entities, we caculate the distances between consumer and fraudster zip codes. Figure 4 (b) provdies a cumulative distribution of the operational radii for cyber and regular fraudsters. This analysis provides insight on weather cyber fraudsters leverage the Internet for more visiblity and access to target more distant individuls. With a median ditance of 993 and 861 for cyber and regular frauds, we believe that both tpyes of fraudsters follow similar trneds. This inidicates that a major the Internet based communicaiton does not provide cyber fraudsters with significant advantage over regular ones as they are able to achieve similar operational spans by using phone and mail based communication methods.

G. Demographic Analysis

We study how prominent demograophic trends such as vary with respect to the fraud categories.

After normalizing the number of complaints in each zipcode against population, we aggregate the complaint rates across medain age, median income, the percentage individuals with

a college degree and the unemployment rate. Figure 5 shows the bar plots of the complaint ratio. We elaborate how frauds these demographics infucence complain the how thinflucence

Age: Over, the ratio of cyber to regular complaints throughout different agr groups remains consistent.

Figure 5 (a) shows tha have a complaint rate of over 50, complain almost twice in comparison to younger indvidials.

Income: We do not observe any significant trends. for low income individuals while more earining individuls are likey to report more cyber frauds Though we do not associte this with targeting we believe this is because more earning individals spend have more access as using the internet more.

Education: While regular fruads remain consistent throughout More educated are more likely to be a vitcim of cyber frauds. We see an increasing ratio ofcyber to regular fraud for an college egucation increases, with higly educated individuals are more likely to be targeret. Similarly, to income we

while regular fraud remains fairly consinstent across different education levers we see that more educated are

Employment: [**?**] Lo

Areas with low unempoyment are likely to complain more...

V. Discussion

Discussion Goes here

VI. Conclusion

The conclusion goes here. Use our methodoloy on more massive datasets. Blablabla said Nobody [1]. [2]

References

[1] N. Jr, "My article," 2006.

[2] A. Smith, L. Rainie, and K. Zickuhr, "College students and technology — pew research center." http://www.pewinternet.org/2011/07/19/college-students-and-technology/, July 2011. (Accessed on 01/17/2017).

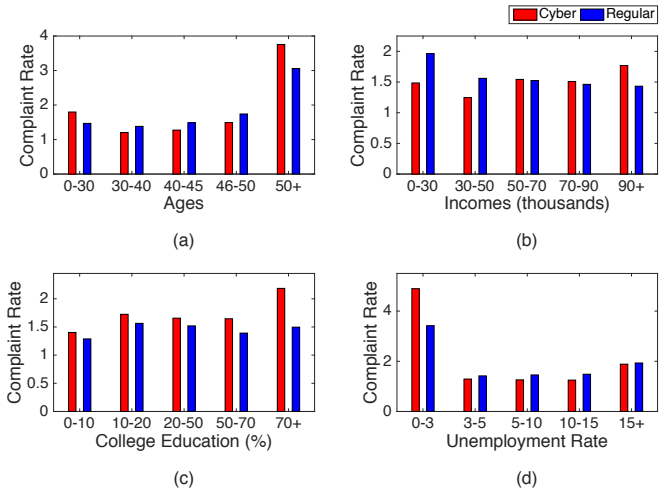

Fig. 5. Cyber and regular fraud trends over age, income, education and unemployment.