Cyber vs Regular: A Comparison of Consumer Fraud in The United States

***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 [13]. 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 [18] . With the number of Internet users on the rise, the number of cyber frauds is likely to increase over the next couple of years [19]. 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 [20]. 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 be targeted. 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 main in from the data followed by our suggestions 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 [13] due to the increased Internet usage trends for sensitive activities. Before the Internet became a primary hub of economic and social activity, researchers measured [9] and developed techniques based on statistical models [10], [11], [12] to detect phone and credit card based frauds. In the past few years, research evaluations have shifted focus

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

towards cyber activity [4], [5], [3], [7] due to its high rate and the potential for harm.

Even though, term *"cyber"* is usually associated with Computer Science, due to its recent socio-economic impact, researchers in Economics, Law, and Finance have also explored for solutions to this problem by incorporating methodologies specific to their areas. Ionescu et. al [3] characterize the types and sources cyber frauds in global digital networks. The authors link the exponential increase of cyber fraud to increased Internet usage for financial management and transactions, especially in financial markets. They suggest the involvement of all stakeholders and employees through awareness and training for containing and mitigating fraud. Similarly, Howard et. al [7] study malicious code attacks against financial networks and suggest technical detection and mitigation techniques for financial infrastructure. [5] 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 [8] 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 [14] uses complaint data to quantify complaint rates and across different ethnic and education groups in the US. [17] also look at how demographics effect the likelyhood of an individual to complain about fraud. Researchers in [4] 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, our work provides a unique angle of evaluation. We provide a comparison for both types of frauds and describe certain complaint charectaristics and demographic trends which more prevalent in each specific category.

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

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 [15]. The specific information that we collect is stated below:

- Population density per zip code
- Employment and income data
- 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 [16]. 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 [15].

B. Calibration Methodology

While information from the US Bureau was standardized and did not require preprocessing, we cater for inconsitencies in the FTC dataset. We realize the the major source of irregularity as a results of its collection ftom various complaint collection agencies furthered by the lack of sanity checks in complaint forms or as a cause of human error while infromation entry. The two major forms of inconsitencies that we filter our are irregular zip codes that cannot be associated with a geographic region in the US and complaints that were missing essential information that required them to be tagged as a cyber of regular fraud.

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

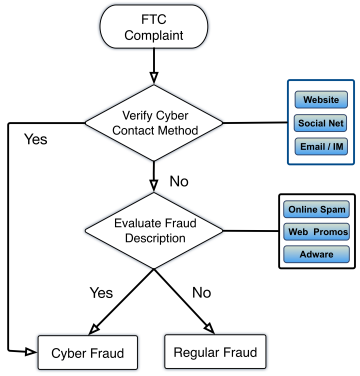

Fig. 1. Classification methodology for cyber and regular frauds

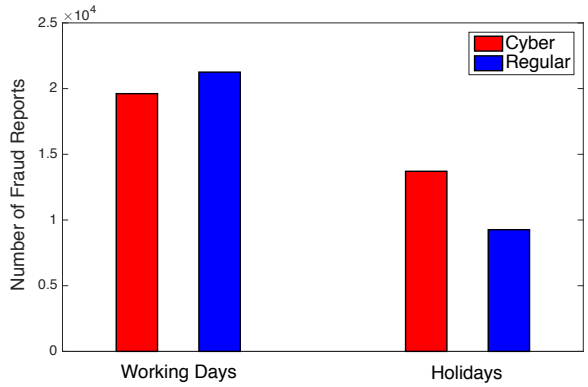

Fig. 2. Cyber and regular fraud variation during regular and holiday season

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 Reporting 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. While the overall rate of reporting 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 days (Aug, 15 to Sept, 5) and holidays (Dec 15, to Jan, 5). Figure 2 shows the variation of cyber and regular frauds within the two specific time periods. We observe that the overall number of fraud reports decrease by a total of 43.7% within the holiday season. Another significant insight is the individual number of cyber and frauds in each time period. While cyber reports are approximately 8.0% less than regular reports during the working days, they are 47.9% more by than regular reports during the holidays.

To further validate our findings from figure 2, we perform a t-test to verify the statistical significance of the reporting difference in the two durations. We evaluate that the *p-value* coefficients by comparing cyber and regular frauds across each region. Table II summarizes our results.

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

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

Our results indicate that cyber fraud reporting trends are statistically significant ($p\text{-value} > 0.05$) between the two time periods. To further investigate phenomenon, we look at the reporting methods vitamins use to contact section IV-B.

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 3 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.

As a major chunk of regular frauds are reported via offline methods, and while offline institutions have reduced operation during holidays, this significantly contributes to the reduction of regular fraud reporting during holidays and biases the value in Figure 2. However, as most cyber frauds are reported via automated online methods which are available around the clock, we believe that cyber frauds experience a true reduction in the holiday period. Our understanding is corroborated by the *p-value* statistics in Table II, where the cyber complaints in the working day and holiday time periods belong to different sets, whereas, the comparisons of corresponding regular complaint sets yield and insignificant value. This insight helps us suggest more meaning suggestions deal with these specific types of crimes in section V

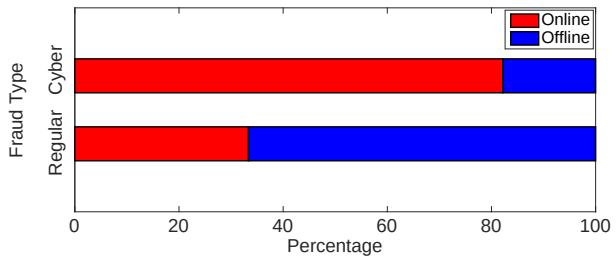

Fig. 3. Distribution of reporting methods in for cyber and regular frauds

C. Consumer Reaction Time

D. Top Fraudsters

Next, we identify the top locations of the fraudsters present in the United States. 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. Table III summarizes the results for the top areas which contribute to a total of 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 III
TOP METROPOLETAN AREAS WHERE FRAUTDSTERS ARE BASED

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 the Buffalo county...

E. Top Frauds

Summary of the top frauds (Maybe include? Not much intutition)

F. Fraud Coverage

Evaluate Distance between Fraud and Comsumer Zip.. See Fraud Coverage...

G. Demographic Analysis

V. DISCUSSION

VI. CONCLUSION

The conclusion goes here. Use our methodololy on more massive datasets.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
[2] pete snyder affiliate fraud
[3] cyberfraud in a global digital network
[4] Consumer Attitudes Toward Data Breach Noti cations and Loss of Personal Information
[5] An Uneven Playing Field: The Advantages of the Cyber Criminal vs. Law Enforcement-and Some Practical
[6] Identity Theft as a Teachable Moment
[7] Cyber Fraud Trends and Mitigation
[8] Cyberfraud in Canada
[9] Controlling Cell Phone Fraud in the US: Lessons for the UK Foresight Prevention Initiative
[10] Neural data mining for credit card fraud detection
[11] Detection of mobile phone fraud using supervised neural networks: A first prototype
[12] Statistical Fraud Detection: A Review
[13] http://gotaclassaction.com/wp-content/uploads/2013/05/FTC-Fraud-Survey.pdf
[14] complaints ftc paper
[15] usbureau website
[16] dept of labor website
[17] Are Consumers Disadvantaged or Vulnerable An Examination of Consumer Complaints to the Better Business Bureau
[18] Link in wunderlist
[19] pew internet study
[20] fbi IC3 Web