An Examination of Co-occurrence of Internet Crimes

Francis Andoh-Baidoo
University of Texas Rio Grande Valley
francis.andohbaidoo@utrgv.edu

Emmanuel Ayaburi
University of Texas Rio Grande Valley
emmanuel.ayaburi@utrgv.edu

Daniel N. Treku
University of Texas Rio Grande Valley
Daniel.treku01@utrgv.edu

Abstract

This paper investigates both the victims' and perpetrators' sides of internet crime to provide an integrated view of the internet crime problem. We seek to understand how the incidences of internet crimes occur across the U.S. states by examining the patterns that exist in internet crime. We collected data from the Federal Bureau of Investigations' internet crime center website. Thirty-eight crime types originally extracted were dimensionally reduced based on crime features and their occurrences. We followed this by factor score cluster analysis. We then examined how the reduced dimensions mapped onto prior literature on crime taxonomy. Based on our analyses, we find that: (1) while some crimes occur together across states, the co-occurrence is not based on neighbor-to-neighbor state ideology; (2) criminal forums is a dominant crime type as it affects over 40 U.S. states and that preventing this crime is key to reducing victim count; (3) there is a significant correlation between many of the crime types identified in this study. Coordinated effort to reduce the effectiveness of criminal forums has the potential to reduce the number of victims. In addition, reducing criminal forums will lead to a reduction in other crimes.

Keywords: Internet crime, Clustering, Criminal forums, Co-occurrence-crime-solution, Law enforcement

1 Introduction

Internet crime, also known as cybercrime, describes the use of computing technology to commit a crime against individuals, organizations, society, and/or properties. Internet crime, with over 38[1] categories, is a concern for citizens, private institutions, law enforcement, and other public institutions. The loss from Internet crimes to a country's economy could be up to 1.5% of its gross domestic product (Lewis, 2018; McAfee, 2014). Internet crime victims suffer huge financial and devastating emotional loss (Modic & Anderson, 2015). In 2017 alone, the Federal Bureau of Investigations (FBI) received a total of 301,580 self-reported complaints from victims of internet crimes, indicating over $1.4 billion losses, which is an increase from the 298,728 complainants in 2016 with a reported loss of more than $1.3 billion, in total, to incidences of cybercrime (Graham, 2017).

The growing trend in cybercrime incidences and associated cost implications continue to motivate researchers to understand the theoretical bases of how and why these crimes are committed (Graham, 2017). To curtail internet crime, several passive solutions from the criminology literature on physical crime have been proofed (e.g., Pratt, Holtfreter, & Reisig, 2010; Yar, 2005). However, internet crime is less constrained by monetary and physical resources and can cause significant harm remotely. Some studies have argued for imposing liabilities on software publishers because of a belief that vulnerabilities in the computer systems aid and abet cybercriminals to commit these internet crimes (e.g., Rusia & Koem, 2005). We note that while such actions may be appropriate in addressing overall cybersecurity challenges, they do not consider the fact that cybercrime could be perpetrated due to or without vulnerabilities in the online computing resources. Awareness is also one such passive mechanism that has vigorously been pursued by law enforcement and governmental agencies to fight the internet crime menace (Burns, Whitworth, & Thompson, 2004). For instance, the FBI cautions U.S. citizens to be wary of becoming victims of internet crimes. Summarily, our review of the cybercrime studies, looking at reducing cybercrime incidences, show that researchers have treated the menace from either a victims' perspective or from perpetrators' perspective (See Table 1A in Appendix A). In assessing the current state in cybersecurity affairs, Sen (2018) argued for a more integrative or holistic approach (i.e., integration of technical, economic, legal, and

[1] https://www.ic3.gov/media/annualreport/2018_IC3Report.pdf

behavioral perspectives) to address the various limitations in existing research works. Following this, our paper looks at the cybercrime menace from a more integrated victim-

perpetrator perspective to understand the nature of the crimes and provide more nuanced insights in dealing with the cybercrime menace.

The FBI has been collecting data through the Internet Crime Complaint Center (IC3) website on diverse internet crimes committed by citizens both within and outside the United States and the associated outcomes on both the victims and subjects. The FBI believes that IC3 affords individuals a reliable and convenient means to report any internet-related criminal activity. The FBI also contends that, through industry partner alliance, law enforcement agencies can use the data for investigative and intelligence purposes and create awareness on Internet-related crimes. However, there is no academic understanding of how the IC3 data contributes to identifying hotspots for optimal prevention and/or reduction of internet crimes. The question worthy of answering is what actionable information can be gleaned from the IC3 data beyond the specific yearly record for each type of crime for each state. We argue that if one were to combine the data on the various internet crimes across the different states, then overall patterns in terms of the natural groupings of the data that goes beyond specific crimes and states might be observed. Such patterns might reveal how different internet crimes are related to one another over the broad set of the U.S. States and how these conditions may explain the outcomes of crimes (victim loss, victim count, subject loss, subject count). Such analyses can help stakeholders as they work in a collaborative environment to mitigate internet crimes. The patterns may reveal the common tools used in internet crimes, the information commonly used or obtained, and the people most affected by these crimes. We argue in this study that an analysis of crime patterns would provide such insights.

Therefore, to address our inquiry into the topography of internet crimes in the U.S., our study uses cluster analysis, a data mining technique, and other statistical analyses to examine the IC3 data set. The results from the cluster analysis add a new angle to the fight against internet crimes by showing which specific crimes or group of crimes a state or a group of states should be concerned about and seek collaboration. For instance, stakeholders can take advantage of the findings to identify unique strategies to address the different clusters identified from the data. We propose law enforcement should be more economical with a one-crime-one-solution approach to cybercrime and adopt a comprehensive co-occurrence crime solution

approach to enjoy significant economies of scale in the fight against internet crime.

2 Relevant Internet Crime Studies

To reduce the likelihood of internet crime, researchers have investigated possible solutions from either the perpetrators' side or victims' side. From the perpetrators' side, it is suggested that increasing certainty and severity of punishment can dissuade perpetrators from engaging in internet crime or complying with security measures (D'Arcy, Hovav, & Galletta, 2009; Herath, Myung-Seong, D'Arcy, Nam, & Rao, 2018). From the victims' side, increasing the level of awareness through education is suggested to be effective at reducing the incidence of successful internet crime (Abbasi, Li, Benjamin, Hu, & Chen, 2014; Hunton, 2009). Researchers (see Table A1) have applied several techniques, including econometric modeling (Hui, Kim, & Wang, 2017; Yue, Wang, & Hui, 2019), text analytics (Abbasi et al., 2014; Samtani, Chinn, Chen, & Nunamaker, 2017; Yue et al., 2019), qualitative review (Holt, 2013; Keyser, 2003; Li, 2007; Yue et al., 2019) and cluster analysis (Abbasi et al., 2014) to examine internet crimes and proffer solutions. Generally, these studies suggest increasing the severity of punishment for those apprehended. However, many cyber-crimes involve professional crime syndicates, which are less sensitive to apprehension and conviction compared to individual offenders (Broadhurst, Grabosky, Alazab, & Chon, 2014; Kshetri, 2010). We argue that solutions rooted in only one-sided views are inadequate for reducing the increasing incidence of internet crime. Reducing the Internet crime problem requires a collaborative effort that aims at identifying the optimal effort to reduce the effect from both the perpetrators' and victims' side, simultaneously. For instance, through strategic alliance and enforcement of shared convention, it is possible to reduce internet crime like denial of service attack (D'Arcy et al., 2009). Likewise, identifying the sources of internet crime, based on their characteristics, their interdependence, and trend simultaneously is critical in combatting cybercrime (Kim, Wang, & Ullrich, 2012).

3 Method and Results

To understand how the incidences of internet crimes occur across the U.S. states, we conducted a quantitative analysis of IC3 data. Thirty-eight crime types were originally extracted into reduced dimensions through exploratory factor analysis using principal component analysis based on crime features and their occurrences. This is followed by factor score cluster analysis using the reduced dimensions and the geographical distribution of the crime types in the U.S. states. We then examined how the reduced

dimensions mapped onto prior literature on crime taxonomy.

3.1 Exploratory Factor Analysis

A two-year panel data set (2016-2017) was obtained from the FBI crime website. The data set consists of information about victim loss, victim count, subject loss, and subject count for each state on the 38 crimes. We analyzed normalized data and extracted its principal components. The Kaiser-Meyin-Oklin measure for sampling adequacy (KMO-MSA) was used to test the suitability of our data set (Cerny & Kaiser, 1977; Kaiser & Rice, 1974). The KMO-MSA values were between 0.83 and 0.95, considering all instances of our component factor extraction, indicating a very good data structure and adequate support for performing PCA. Principal component analysis was used to transform the high dimensions of crime types to a low dimensional space (Spicer, 2005) for each outcome (see Table A2 in Appendix). The criteria we used to determine the number of factors was eigenvalue > 1.

We categorized the components of the PCA results for each of the internet crime outcomes based on their descriptions as follows: victim count (victim-perpetrator familiarity vs. non-victim-perpetrator familiarity), victim count (platform vs. non-platform based), subject loss (exploitative vs. non-exploitative), and subject count (nation-state vs. non-nation state). These categories map onto Howard's (1997) framework. This framework presents a theoretical taxonomy for classifying Internet security attacks. We argue that such a framework provides a theory grounded approach to classify internet crimes. The taxonomy consists of five variables: attacker (hackers, spies, terrorists, corporate raiders, professional criminals, or vandals), tools (user command, script or program, autonomous agent, toolkit, distributed tool, or data taper), access (unauthorized use or unauthorized access), results (corruption of information, disclosure of information, theft of service, or DoS) and attacker objective (challenge/status, political gain, financial gain, and damage).

This framework has been applied to study diverse information security issues such as information systems threats (Im & Baskerville, 2005) and internet security breaches (Andoh-Baidoo & Osei-Bryson, 2007). In this work, the results of Howard's framework maps onto the outcome of the internet crime data used in the cluster analysis. As shown in Table A2, four categories of Howard's framework relate to the crime categories of the four internet crime outcomes – access type (victim loss), tools used (victim count), objectives (subject-loss), an attacker type (subject count). This taxonomy helps us understand the tools used, the information weaponized and the people targeted by perpetrators of internet crime.

3.2 Cluster Analysis

We employed cluster analysis, an unsupervised learning data mining method (Jain, Murty, & Flynn, 2000), as a robust method to determine homogenous groups of states that are susceptible to a set of crime (extracted factors) (Jain & Dubes, 1998). K-means clustering, a prototype-based and partitional clustering technique, is the widest cluster algorithm employed and can be used for various data types such as documents and time series. K-means clustering algorithm attempts to find a user-specified number of clusters (K) represented by centroids or cluster centers (Tan, Steinbach, Karpatne, & Kumar, 2019). One of the critiques leveled against the K-means clustering algorithm is its poor handling of certain types of data structure such as spherical data or very sparse data, as seen in most document analyses (Hornik, Feinerer, Kober, & Buchta, 2012; Madhulatha, 2012). Our balanced panel data, however, is more rectangular in structure and suits the measurement of distances between two objects in the Euclidean space. Further, several proximity functions such as cosine and Bregman divergence can be used to test the robustness of centroids or final cluster centers with this approach. K-means approach also works well with small to large datasets (Madhulatha, 2012).

Since the K-means technique requires that the number of clusters to generate be specified before running the algorithm, we first used the two-step clustering algorithm to determine the number of clusters to extract. The K-means algorithm was then used to iteratively estimate the cluster means and assign each case to the cluster for which its distance to the cluster mean is the smallest (Hartigan & Wong, 1979; Tan et al., 2019). Data visualization was used to check outliers that may unduly influence the clusters. During cluster center assignments, the squared Euclidean space k-means proximity method chooses the means based on the assignments which produce the minimum sum of squared error. Our estimation is underpinned by this method.

Given a collection of data points and a parameter k, we find k centroids in a d-dimensional space by estimating the mean of the centroids that minimize the sum of squared error (SSE). In one-dimensional space, the mean of the centroid, which minimizes the SSE, is given by:

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} dist(c_i, x)^2$$

$$= \sum_{i=1}^{K} \sum_{x \in C_i} (c_i - x)^2$$

Where *Ci* is the i^{th} cluster, *x* is a data point in C_i and c_i in *dist (ci, x)* is the mean of the i^{th} cluster. By

differentiating the one-dimensional space equation, we can solve for the k^{th} centroid, C_k that minimizes SSE in a d-dimensional space, as shown below:

$$\frac{\partial SSE}{\partial c_k} = \frac{\partial}{\partial c_k} \sum_{i=1}^{K} \sum_{x \in C_i} (c_i - x)^2$$

$$\frac{\partial SSE}{\partial c_k} = \sum_{i=1}^{K} \sum_{x \in C_i} \frac{\partial}{\partial c_k} (c_i - x)^2$$

$$= \sum_{x \in C_i} 2 \, x \, (c_k - x_k) = 0$$

$$m_k c_k = \sum_{x \in C_k} X_k$$

$$C_k = \frac{1}{m_k} \sum_{x \in C_k} X_k$$

This solution best updates a cluster centroid in the k-means iterative processes. Thus, the optimized solution stated above informed the final cluster centers used in our analyses of the four internet crime outcomes: victim loss, victim count, subject loss, and subject count.

3.2.1 Victim Loss

From our analysis, the mean and F values are (24.065, f (2, 58) =149.204, p<0.01) and (21.677, f(2, 58) =87.373, p<0.01) for non-victim-perpetrator familiarity and victim-perpetrator familiarity crime categories, respectively. This suggests that the level of victim-perpetrator familiarity is a major contributor to the amount of loss. Figure 1 and Table 1 show that cluster 1, which consists of only California, has an average victim loss much higher than clusters 2 and 3 for victim-perpetrator familiarity-based crimes.

Table 1. Final Cluster Centers

Crime Types	**Cluster**		
	1	2	3
Victim-Perpetrator familiarity	**6.87**	.09	**-.13**
Non-victim-Perpetrator familiarity	-.54	**4.57**	-.16
States	CA	FL, NY	AL, AK, AS, AZ, AR, CO, CT, DE, DC, GA, GU, HI, ID, IL, IN, I.O., KS, KY, LA, ME, MD, MA, MI, MN, MS, MT, MO, NE, NV, NH, NJ, NM, NY, NC, ND, MP, OH, OK, OR, PA, PR, RI, SC, SD, TN, TX, ISO, UT, VT, VI, VA, WA, WV, WI, WY

Figure 1. Victim Loss Clusters by States-based Crime Types

Cluster 2, consisting of Florida and New York, has an average victim loss that is much higher than clusters 1 and 3 with respect to the crimes in non-victim-perpetrator familiarity. Cluster 3, represented by over 50 states, has below average victim loss for all crime types. In order words, from the cluster analysis regarding victim loss resulting from cybercrime, California must pay attention to victim-perpetrator familiarity crimes, Florida and New York must pay more attention to non-victim-perpetrator familiarity crimes.

3.2.2 Victim Count

The mean and F values are (8.355, f(2, 58) =87.623, p<0.01) and (15.959, f(2, 58) =156.564, p<0.01) for non-platform-based and platform-based respectively, suggesting that platform-based contributes more to the separation of the clusters than non-platform based crimes. This suggests that the tools employed by the perpetrators are key to understanding victim count. In this study, the results indicate that the criminal forum platform, where perpetrators share and gain knowledge on internet crime, is an important mechanism that requires the attention of internet crime prevention stakeholders. Figure 2 and Table 2 illustrate cluster 1, which consists of 19 states, has an average victim count for crimes in non-platform based and below-average victim count for crimes in platform-based. Cluster 2, with a much larger number of states, has below average victim count for all crime types. Cluster 3, consisting of California, Florida

, and New York, has an average victim count that is much higher than clusters 1 and 2 with respect to platform-based crimes. Thus, victim count can be reduced by reducing criminal forums in California, Florida, and New York.

Table 2. Final Cluster Centers

	Cluster		
Crime Types	1	2	3
Non-platform-based	**0.45**	-.24	**-1.9**
Platform-based	-.64	**-.03**	**2.9**

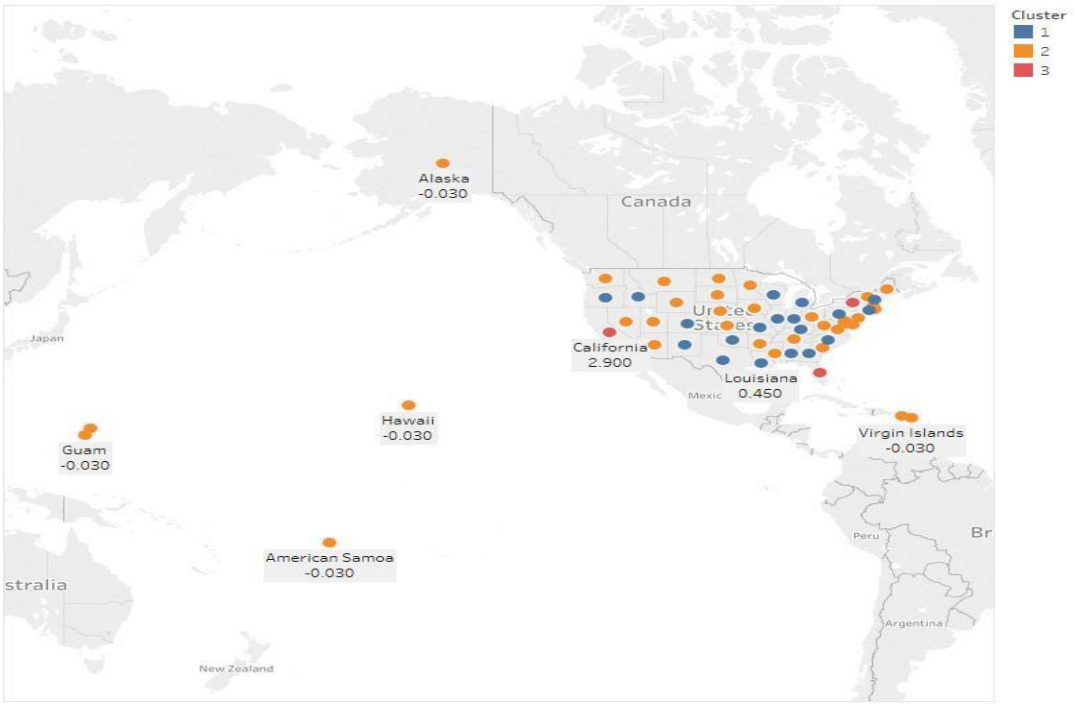

Figure 2. Victim Count Clusters by States-based on Crime Types

3.2.3 Subject Loss

The mean and F values are (5.595, f (1, 58) = 6.106, p<0.05) and (31.523, f(1, 58) = 70.834, p<0.01) for non-exploitative and exploitative crime types, respectively, suggesting that exploitative crimes contribute more to the separation of the clusters than non-exploitative. Figure 3 and Table 3 depict cluster 1, which consists of all states except Florida, Georgia, and Texas has below average victim count for all crime types. Cluster 2, which consists of Florida, Georgia, and Texas, has an average victim count that is much higher than cluster 1 with respect to the crimes in exploitative. This suggests that care should be taken to address

crimes that exploit the vulnerable, especially children.

Table 3. Final luster Centers

	Cluster	
Crime Types	1	2
Non-Exploitative	**-.07**	1.33
Exploitative	-.18	**3.12**
States	AL, AK, AS, AZ, AR, CA, CO, CT, DE, DC, GU, HI, ID, IL, IN, IA, KS, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NE, NV, NH, NJ, NM, NY, NC, ND, Northern Mariana Islands (M.P.), O.H., OK, OR, PA, PR, RI, SC, SD, TN, United States Minor Outly, UT, VT, VI, VA, WA, WV, WI, WY	FL, GA, TX

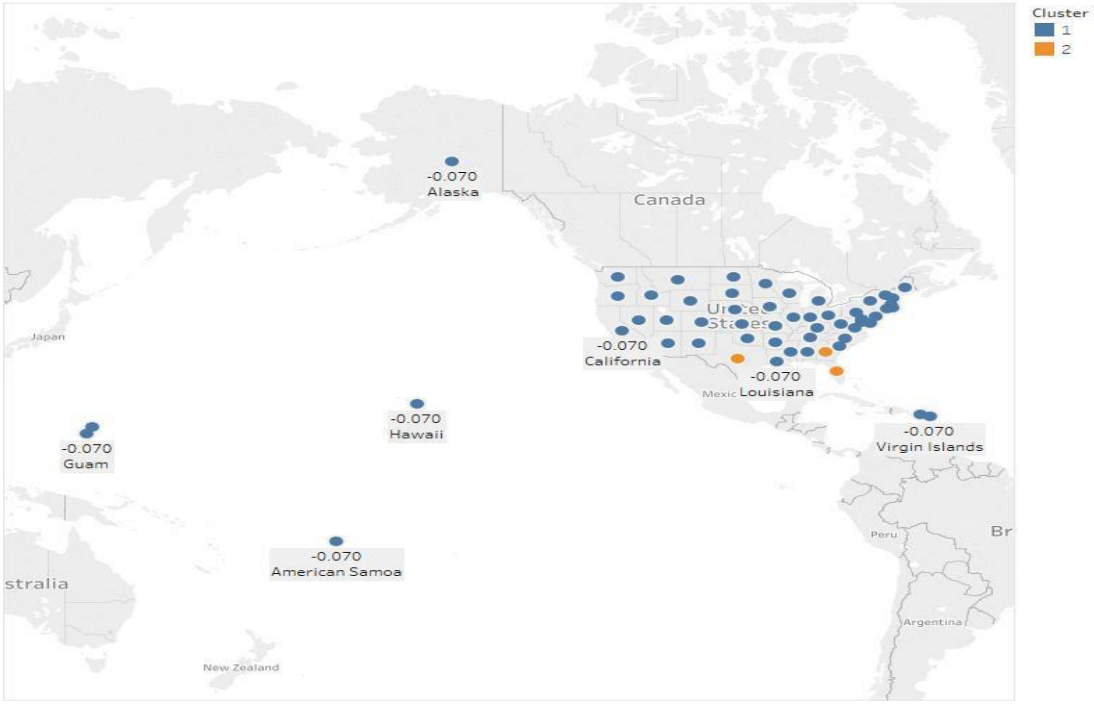

Figure 3. Subject Loss Clusters by States-based on Crime Types

3.2.4 Subject Count

The mean and F values are (16.487, f (2, 58) = 38.667, p<0.01) and (21.882, f(2, 58) = 96.565, p<0.01) for non-nation state and nation-state crimes respectively suggesting that nation-state contributes more to the separation of the clusters. Figure 4 and Table 4 display that cluster 1, which consists of only District of Columbia and New York, has an above-average subject count for crimes in nation-state and below-average victim count for non-nation state crimes. Cluster 2, which consists of all the states except the District

of Columbia and New York, has an above-average victim count for all crime types. However, the subject count is higher for the non-nation-state than the nation-state, and the number of crimes in the nation-state is just three compared to 35 for the non-nation state. Thus, while New York and the District of Columbia can focus on crimes in nation-state to reduce subject count, all the other states must address all crimes.

Table 4. Final Cluster Centers

	Cluster	
Crime Types	1	2
Non-nation State	-1.34	**5.4**
Nation State	**4.4**	1.72
States	DC, NY	AL, AK, AS, AZ, AR, CA, CO, CT, DE, FL, GA, GU, HI, ID, IL, IN, IA, KS, KY, LA, ME, MD, MA, MI, MN, MS, MO, MT, NE, NV, NH, NJ, NM, NC, ND, Northern Mariana Islands, OH, OK, OR, PA, PR, RI, SC, SD, TN, TX, United States Minor Outly, UT, VI, VA, WA, WV, WI, WY

Figure 4. Subject Count Clusters by States-based on Crime Types

4 Discussion

For law enforcement to reduce or eliminate internet crimes, it is ideal that optimal effort is made in multiple fronts simultaneously. We have extended the application of Howard's (1997) taxonomy to classify various internet crimes across states in the U.S. As Law enforcement apply this taxonomy, they can better understand the related characteristics of internet crimes. Our principal component analysis and cluster analysis reveal interesting results. As shown in Figure 1, for victim loss, the clustering analysis suggests that Florida and New York must deal with the following crimes (civil matter, identity theft, pr_copyright and counterfeit, other crimes, phishing_vishing smishing pharming, terrorism, and virtual currency) while California deals with all the other crimes. Increase awareness/education of people may be an appropriate solution for addressing victim-perpetrator familiarity crimes. In contrast, non-victim perpetrator crimes may require direct intervention using technical solutions such as installing protective information technologies. Regarding victim count, the clustering analysis suggests that one crime, i.e., criminal forum determines the size of victim count for 38 states. More importantly, except California, Florida, and New York, the effect on the other 35 states is negligible. Reducing the count of the victims in these 38 states by targeting the tools used will lower the victim loss tremendously. Where more effort must be made are the following states; A.L., CO, CT, GA, ID, IL, IN, KY, LA, MI, MO, NH, NM, NC, OK, OR, PA, TX, WI, where
the victim count depends on all the crimes except criminal forum. One such effort is for law enforcement to understudy criminal forums, obtain sensitive information, and to share preventive and reactive solutions.

Subject loss presents the most straightforward problem, although it is of the least concern among the four internet crime outcomes. Issues related to victims are of uttermost importance. Subject loss is easiest to solve for several reasons. First, there are only two clusters. Second, only one cluster involving three states poses threats. Third, crime propensity is higher in exploitative crime categories, consisting of only two crimes. This category is over three times higher than non-exploitative, which consists of the rest of the U.S. states and territories. Reducing exploitative-based crimes can be achieved through education and awareness programs for the most vulnerable and may lead to the adoption or use of protective technologies/techniques (Chatterjee, Arpan Kumar, Dwivedi, & Kizgin, 2018)

Subject count is an important problem to resolve because reducing the number of subjects (criminals) involved in the crime can reduce the effect of the crime on the victims. An effective solution should address

the following crimes: Government impersonation, hacktivist, extortion, gambling in District of Columbia, and New York, and all other crimes in all the other states. Nation-state crimes require collaboration among law enforcement agencies and other state institutions across states.

Concerning the victim outcomes, while the magnitude of loss related to the two factors is higher than those of count, it seems that it may be easier to resolve the loss problem. Cluster 3, with over 50 states, does not present a major problem. Cluster 1 with the highest crime propensity involves about 29 crimes in California, while cluster 2 with 9 crimes in Florida and New York. Reducing victim count requires addressing criminal forums targeted at three states- California, Florida, and New York while curtailing all crimes except criminal forums in 19 states.

For clusters that involve one single state, the correct solution may be offered by that state alone. Researchers argue that while it may be true that domestic law enforcement agencies have not deterred cybercrimes, their actions against security violations might be responsible for why attackers are initiating attacks from other nations (D'Arcy et al., 2009; Hunton, 2009). Thus, in the same way, an effective solution provided by the state will cause criminals to look elsewhere, which could be other states or outside the U.S. However, in this case, the results suggest that other states may not suffer threats that the specific states face. Therefore, it would be difficult for cybercriminals to find a suitable geographic target, at least in the U.S. states and territories. For clusters that involve several states, providing a good solution demand collaborative effort across law enforcement and other government institutions. The call is due to the difficulty in the processes involved in collaborative enforcement such as identifying and tracking cybercriminals, assessing the extent and impact of offenses, and the collection and analysis of digital evidence related to the crime (Hui et al., 2017; Png, Wang, & Wang, 2008).

Some studies have reported negative consequences of online information exchange (e.g., Hunton, 2009). However, recent studies indicate that discussions on criminal forums may exhibit dual-use characteristics (Yue et al., 2019). In a study of hacker forums, it is argued that "hacking discussion may contribute to developing and spreading of protection knowledge" [p.74]. In addition, "online discussion of distributed denial of service (DDOS) attacks in hackforums.net decreases the number of DDOS-attack victims" [p.73].

Thus, we contend that identifying solutions for criminal forums could address victim count problem because

criminal forums are the only crime responsible for the high victim count in California, Florida, and New York. Given that criminal forums are relatively higher than all the crimes that affect additional 19 states and that none of the crimes affect the remaining 40 states, an appropriate solution is needed to address criminal forums. Thus, gaining knowledge from the criminal forums is an appropriate means for citizens to become aware of the sophistication of internet crimes, tools, and techniques that they can use to protect themselves and increase awareness in heightened internet crime activities. At the minimum, law enforcement and other relevant state agencies can understudy in such criminal forums to gain access to the knowledge that the criminals on such forums have built over the years and disseminate such knowledge to the general public.

5 Implications

From our analysis of self-reported internet crime data collected by the US FBI, the findings provide insights about cybercrimes against society, property, individuals, and organizations. For societies, the promotion of mutual cooperation between law enforcement agencies across the different states is key to tackling cybercrime. This collaboration may not be based on neighbor-to-neighbor state ideology. States that should collaborate according to the cluster results do not necessarily share physical geographical boundaries. This is because internet crimes such as criminal forums that are popular in California, New York, and Florida involve professional crime syndicates (Kshetri, 2010) and require concerted effort to be taken down. Enforcement efforts that have spillover effects as various cybercrime reduction techniques are complementary (D'Arcy et al., 2009). We provide pioneering evidence when law enforcement across these states make the optimal investment to reduce criminal forums, the number of individuals (victim count) would be greatly reduced (Figure 2), and this may lead to lower amount loss (Figure 1). For individuals on the victims' side, when increased awareness is created through enhanced cybersecurity education, it may prevent individuals from becoming victims, the amount of loss would be greatly reduced. A news report indicated that a single crime (e.g., romance scam) could involve several victims and subjects across states (i.e., New Jersey, Alaska, Oklahoma, Florida, Texas, Kansas, and Iowa) and nations (Buono, 2014). Thus, most agencies tasked with internet crime reduction need to spend resources wisely to ensure continuous funding from their respective states. These resources could be put to better use if there were no cybercrime. In fighting certain crimes, these agencies can scale up in the long run when they identify which states to

collaborate. For example, it would be optimal for states in Florida, Georgia, and Texas to collaborate to fight crimes against children, malware, scareware, and virus effectively (see Figure 3). One such approach might include intensification of the education of parents who play a critical role in protecting the children from becoming victims of cybercrime (Tennakoon, Saridakis, & Anne-Marie, 2018). Many cybercrime-related opportunity costs, such as the procurement and operation of protection technologies and mandated security audits and contingency planning, are incurred on an ongoing basis. For organizations, given the significant correlation between many of the crime types (Principal component analysis results) identified in this study, we propose law enforcement should be more economical with a one-crime-one-solution approach to cybercrime but adopt a comprehensive co-occurrence crime solution approach. This will ensure that law enforcement organizations enjoy significant economies of scale in the fight against internet crime.

6 Conclusions

We have shown how data analytic tools are critical in enabling coordination between stakeholders such as law enforcement, policymakers, and funding agencies, to reduce the occurrence of internet crime and associated effects. Recently, a syndicate group of individuals located in several countries and across several states was arrested for specific crimes. This suggests that subjects may work together to perpetrate a group of crimes across a geographical region. Nevertheless, it is acknowledged that the numerous stakeholders involved in the fight against internet crime should coordinate to maximize their limited resources. To enjoy economies of scale, agencies should invest these resources in a targeted fashion that results in optimal crime reduction. Our results, for example, show that coordinated effort to reduce the effectiveness of criminal forums by states such as California, New York, and Texas, has the potential to reduce the victims of internet crime. In addition, a mechanism for reducing criminal forums will lead to a reduction of other crimes, such as Advanced fee or confidence fraud romance based on the loss amount, as they have a high probability of co-occurrence. Further research should look towards a more longitudinal dataset, as it would be interesting to find out how the occurrence or the amount of victim loss change over time to identify if there is seasonality in the occurrence of these crimes.

7 Limitations and Future Research

One of the limitations of the study is that the website presents data for only a two-year period. The data provided is at the aggregate level making it difficult to perform analysis on the individual incidences.

Hopefully, the FBI will continue to collect more data such that future research could perform longitudinal analysis.

Future research could combine FBI data with other state data to investigate why certain locations suffer more specific crime categories. For instance, why is California more susceptible to victim-perpetrator familiarity, but New York and Florida face non-victim-perpetrator familiarity crime threats? Similarly, why do those three states have a high propensity to criminal forums? Ineffective cognitive processing has been argued as a key reason for the victimization of individuals by cybercriminals. Therefore, Vishwanath, Harrison, & Ng (2018) proposed the suspicion, cognition, automaticity model (SCAM) to help explain how "spearphishing" (highly personalized human phishing) attacks take advantage of people's weaknesses in online behavior (Vishwanath et al., 2018). Concerning our findings, future studies may look at how SCAM explains people's weaknesses for co-occurrence crimes or how the interrogation of SCAM within the boundaries of our victim-perpetrator crime categorizations can enhance the efficiency in the cybercrime reduction efforts.

References

Abbasi, A., Li, W., Benjamin, V., Hu, S., & Chen, H. (2014). Descriptive Analytics: Examining Expert Hackers in Web Forums. *2014 IEEE Joint Intelligence and Security Informatics Conference*, 56–63. https://doi.org/10.1109/JISIC.2014.18

Andoh-Baidoo, F. K., & Osei-Bryson, K. M. (2007). Exploring the Characteristics of Internet Security Breaches that Impact the Market Value of Breached Firms. *Expert Systems with Applications*, *32*(3), 703–725.

Benjamin, V., Zhang, B., Nunamaker, J. F., & Chen, H. (2016). Examining Hacker Participation Length in Cybercriminal Internet-Relay-Chat Communities. *Journal of Management Information Systems*, *33*(2), 482–510. https://doi.org/10.1080/07421222.2016.1205918

Broadhurst, R., Grabosky, P., Alazab, M., & Chon, S. (2014). Organizations and Cybercrime: An Analysis of the Nature of Groups engaged in Cyber Crime. *International Journal of Cyber Criminology*, *8*(1), 1–20.

Buono, L. (2014). Fighting Cybercrime Through Prevention, Outreach, and Awareness Raising. *ERA Forum*, *15*(1), 1–8. https://doi.org/10.1007/s12027-014-0333-4

Burns, R. G., Whitworth, K. H., & Thompson, C. Y. (2004). Assessing law enforcement preparedness to address Internet fraud. *Journal of Criminal Justice*, *32*(5), 477–493. https://doi.org/10.1016/j.jcrimjus.2004.06.008

Cerny, C. A., & Kaiser, H. F. (1977). A Study of a Measure of Sampling Adequacy for Factor-Analytic Correlation Matrices. *Multivariate Behavioral Research*, *12*(1), 43–47.

Chatterjee, S., Arpan Kumar, K., Dwivedi, Y. K., & Kizgin, H. (2018). Prevention of Cybercrimes in Smart Cities of India: From a Citizen's Perspective. *Information Technology & People*, *ahead-of-print*.

Clough, J. (2012). The Council of Europe Convention on Cybercrime: Defining `Crime' in a Digital World. *Criminal Law Forum*, *23*(4), 363–391. https://doi.org/10.1007/s10609-012-9183-3

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User Awareness of Security Countermeasures and its Impact

on Information Systems Misuse: A Deterrence Approach. *Information Systems Research*, *20*(1), 79–98.

GhanaWeb. (2018, November 7). US-based Ghanaian arrested for $5m romance fraud. Retrieved April 22, 2019, from Crime & Punishment website: https://www.ghanaweb.com/GhanaHomePage/NewsArchive/US-based-Ghanaian-arrested-for-5m-romance-fraud-698762#

Graham, R. S. (2017, October 18). The Difference Between Cybersecurity and Cybercrime, and Why It Matters [Research Report]. Retrieved October 3, 2019, from The Conversion website: http://theconversation.com/the-difference-between-cybersecurity-and-cybercrime-and-why-it-matters-85654

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*, 100–108.

Herath, T., Myung-Seong, Y., D'Arcy, J., Nam, K., & Rao, H. R. (2018). Examining Employee Security Violations: Moral Disengagement and Its Environmental Influences. *Information Technology & People*, *31*(6), 1135–1162.

Holt, T. J. (2013). Examining the Forces Shaping Cybercrime Markets Online. *Social Science Computer Review*, *31*(2), 165–177. https://doi.org/10.1177/0894439312452998

Hornik, K., Feinerer, I., Kober, M., & Buchta, C. (2012). Spherical k-Means Clustering. *Journal of Statistical Software*, *50*(10), 22.

Howard. (1997). *An Analysis of Security Incidents on the Internet 1989—1995* (PhD Thesis, Carnegie Mellon University). Retrieved from https://resources.sei.cmu.edu/asset_files/WhitePaper/1997_019_001_52455.pdf

Hui, K.-L., Kim, S. H., & Wang, Q.-H. (2017). Cybercrime Deterrence and International Legislation: Evidence from Distributed Denial of Service Attacks. *MIS Quarterly*, *41*(2), 497–523. https://doi.org/10.25300/MISQ/2017/41.2.08

Hunton, P. (2009). The Growing Phenomenon of Crime and the Internet: A Cybercrime Execution and Analysis Model. *Computer Law & Security Review*, *25*(6), 528–535. https://doi.org/10.1016/j.clsr.2009.09.005

Im, G. P., & Baskerville, R. L. (2005). A Longitudinal Study of Information System Threat Categories: The Enduring Problem of Human Error. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *36*(4), 68–79.

Jain, A. K., & Dubes, R. C. (1998). Algorithms for clustering Data. *Prentice-Hall, Inc. Upper Saddle River, NJ,*.

Jain, A. K., Murty, M. N., & Flynn, P. J. (2000). *Data Clustering: A Review*. IEEE Computer Society Press.

Kaiser, H. F., & Rice, J. (1974). KMO in SPSS (Kaiser and Rice, 1974)—, 1974 vol 34, pages 111-117. *Educational and Psychological Measurement*, *34*, 111–117.

Keyser, M. (2003). The Council of Europe Convention on Cybercrime. *Journal of Transnational Law and Policy*, *12*(2), 287–326.

Kim, S. H., Wang, Q.-H., & Ullrich, J. B. (2012). A Comparative Study of Cyberattacks. *Communications of the ACM*, *55*(3), 66. https://doi.org/10.1145/2093548.2093568

Kshetri, N. (2010). Diffusion and Effects of Cyber-Crime in Developing Economies. *Third World Quarterly*, *31*(7), 1057–1079. https://doi.org/10.1080/01436597.2010.518752

Lewis, J. (2018). *Economic Impact of Cybercrime—No Slowing Down Report* (p. 28). Retrieved from McAfee and Center for Strategic and International Studies website: https://csis-prod.s3.amazonaws.com/s3fs-public/publication/economic-impact-cybercrime.pdf

Li, X. (2007). International Actions against Cybercrime: Networking Legal Systems in the Networked Crime Scene. *Webology*, *4*(3), 13.

Madhulatha, T. S. (2012). An Overview on Clustering Methods. *IOSR Journal of Engineering*, *2*(4), 719–725. https://doi.org/10.9790/3021-0204719725

McAfee. (2014). *Net Losses: Estimating the Global Cost of Cybercrime*. Retrieved from McAfee and Center for Strategic and International Studies website: https://csis-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/attachments/140609_rp_economic_impact_cybercrime_report.pdf

Modic, D., & Anderson, R. (2015). It's All Over but the Crying: The Emotional and Financial Impact of Internet Fraud. *IEEE Security & Privacy*, *13*(5), 99–103. https://doi.org/10.1109/MSP.2015.107

Png, I. P. L., Wang, C. Y., & Wang, Q. H. T. (2008). The Deterrent and Displacement Effects of Information Security Enforcement: International Evidence. *Journal of Management Information Systems*, *25*(2), 125–144.

Png, I. P. L., & Wang, Q.-H. (2009). Information Security: Facilitating User Precautions Vis-à-Vis Enforcement Against Attackers. *Journal of Management Information Systems*, *26*(2), 97–121. https://doi.org/10.2753/MIS0742-1222260205

Pratt, T. C., Holtfreter, K., & Reisig, M. D. (2010). Routine Online Activity and Internet Fraud Targeting: Extending the Generality of Routine Activity Theory. *Journal of Research in Crime and Delinquency*, *47*(3), 267–296. https://doi.org/10.1177/0022427810365903

Rusia, M. L., & Koem, T. H. (2005). The Tort of Negligent Enablement of Cybercrime. *Berkeley Technology Law Journal*, *20*(4), 1553–1611.

Samtani, S., Chinn, R., Chen, H., & Nunamaker, J. F. (2017). Exploring Emerging Hacker Assets and Key Hackers for Proactive Cyber Threat Intelligence. *Journal of Management Information Systems*, *34*(4), 1023–1053. https://doi.org/10.1080/07421222.2017.1394049

Sen, R. (2018). Challenges to Cybersecurity: Current State of Affairs. *Communications of the Association for Information Systems*, *43*(2), 22–44. https://doi.org/10.17705/1CAIS.04302

Spicer, J. (2005). *Making Sense of Multivariate Data Analysis: An Intuitive Approach* (6th ed.). California: SAGE Publications Inc.

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). *Introduction to Data Mining* (2nd ed.). New York, NY: Pearson Education.

Tennakoon, H., Saridakis, G., & Anne-Marie, M. (2018). Child Online Safety and Parental Intervention: A Study of Sri Lankan Internet Users. *Information Technology & People*, *31*(3), 770–790.

Vishwanath, A., Harrison, B., & Ng, Y. J. (2018). Suspicion, Cognition, and Automaticity Model of Phishing Susceptibility. *Communication Research*, *45*(8), 1146–1166. https://doi.org/10.1177/0093650215627483

Yar, M. (2005). The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory. *European Journal of Criminology*, *2*(4), 407–427. https://doi.org/10.1177/147737080556056

Yue, W. T., Wang, Q.-H., & Hui, K.-L. (2019). See No Evil, Hear No Evil? Dissecting the Impact of Online Hacker Forums. *MIS Quarterly*, *43*(1), 73–95. https://doi.org/10.25300/MISQ/2019/1304

Appendix A: Cybercrime Literature and Cybercrime Types

Table A1. Summary of Some Key Studies on Internet Crime

Study	Theoretical Lens and Methodology	Findings
Impact of Online Hacker Forums (Yue et al., 2019)	Exploring the impact of online channels on offline outcomes regarding the dual-use nature of online hacking and moral ambiguity of hacking. Empirical insight using text analytics and econometric modeling of distributed denial of service (DDOS) online discussion topics	Discussion topics with similar keywords can variously increase or decrease DDOS attacks. Increase in discussion decreases DDOS attacks. Mentioning botnets, especially new botnets, increases attacks, but follow-up discussions decrease the attacks. An online-hacker-forum discussion could have both negative and positive consequences given the same discussion topic.
Cybercrime deterrence and international legislation (Hui et al., 2017)	Examination of the effect of Deterrence Enforcement and punishment at the country level using econometric modeling	Enforcing the Convention on Cybercrime (COC) in deterring distributed denial of service (DDOS) attacks decreases the attacks by at least 11.8 percent. A similar deterrence effect does not exist if the enforcing countries make reservations on international cooperation. Cyber attackers can be motivated by economic incentives and are strategic in choosing attack targets. They respond to heightened law enforcement by either forgoing their attacks or shifting attacks to non-enforcing countries.
COC and the definition of digital crime (Clough, 2012)	A critical qualitative review of the Convention on Cybercrime (COC)	The Convention on Cybercrime (COC) is not a model law, but a framework upon which specific offenses can be based. It allows countries to modify their laws, where necessary, to keep pace with technology. Although the provisions of the convention are imperfect, they remain largely relevant today. Mechanisms for improvement on cybercrime prevention are built into the convention.
Internationati-onal impetus of combatting cybercrime(Li, 2007)	Review on COC and other multinational efforts (U.N., Commonwealth) on cybercrime.	Actions of international harmonization are classified into professional, regional, multinational and global actions COC has an influence on the state (country) and international levels of legal cybercrime countermeasure (preventive measures).
European Convention on Cybercrime (Keyser, 2003)	Critical literature review on COC	COC Report on crime prevention.

User Precautions Vis-a-Vis Enforcement (Png & Wang, 2009)	Economics of information security and optimization of user behavior	For both mass and targeted attacks, facilitating end-user precautions reduces the expected loss of end-users The impact of enforcement on expected loss depends on the balance between deterrence and slackening of end-user precautions. With targeted attacks, facilitating end-users' [victims] precautions are more effective for users with a relatively high valuation of information security, while enforcement against attackers [subjects] is more effective for users with a relatively low valuation of security.
Hacker participation in Internet relay chat (IRC) communities (Benjamin, Zhang, Nunamaker, & Chen, 2016)	IRC architecture using text analytics	Specific Internet Relay Chat cybercriminal community behaviors or features are unique to long-term participants and less attributed to shorter-term participants. Distinct in-degree and out-degree ties are key in the cybercommunity. Participants who create many distinct ties are characterized by longer periods of active participation. Content (texts) features were not significant.
Proactive cyber threat intelligence (Samtani et al., 2017)	Understanding Cyber threat intelligence using text mining and social network analysis	Tools such as crypters, keyloggers, web, and database exploits may have been the cause of recent breaches against organizations such as the Office of Personnel Management (OPM).
Analytics of Expert Hackers' Forums (Abbasi et al., 2014)	Online hacker social dynamics using text analytics	The study proposed a social media analytical model that can be applied to various forms of user-generated content (UGC) by analyzing both structural features and content features. The study provides a complete analytical framework to analyze the key hackers from both the interaction network and discussion content perspectives. The framework can benefit cybersecurity researchers and practitioners by offering an inclusive angle for analyzing hackers' social dynamics. Cluster analysis identified four types of hacker forums users
Cybercrime markets (Holt, 2013)	Analyses of 909 Russian fora (cybercommunity) threads using a grounded theory approach	Price, customer service, and trust influence: (i) the relationships between actors in malware and hacking markets (black market actors), and (ii) the nature of exchanges in these cybercrime forums.
Comparative study of cyberattacks (Kim et al., 2012)	Information security as technical, business and critical policy issue through analysis of country-level data. Information security externalities	Three lessons are advanced in combatting cybercrime: (i)Identify the top sources of attacks from demographic characteristics, (ii) Global diffusion trend across regions, not countries, and (iii)Considerable interdependence of global trends and compelling substitution effect. Factor analysis found that of the top 16 countries ever listed as a top-10 country for attack origin (2005–

		2009), 49% of attacks could be explained by a single (general) factor that can be labeled 'global co-movement'. Four steps are advanced for consideration by countries along – measurement, responsibility, collaboration, and constitutional conflict.

Table A2. Crime Categorization

Outcome	**Crime Types**	**Crime Category**	**Reference to Howard's framework (Howard, 1997)**
Victim loss	Advanced fee, bec eac, charity, confidence fraud romance, corporate data breach, credit card fraud, crimes against children, criminal forums, denial of service, employment, extortion, gambling, government impersonation, hacktivist, harassment threats of violence, investment, lottery sweepstakes inheritance, malware scareware virus, misrepresentation, nonpayment non delivery, overpayment, personal data breach	Victim-Perpetrator familiarity	Access
	Civil matter, identity theft, pr_copyright and counterfeit, other crimes, phishing vishing smishing pharming, terrorism, virtual currency	Non-victim-Perpetrator familiarity	
Victim Count	advanced_fee, bec_eac, charity,confidence_fraud_romance, corporate_data_breach, credit_card_fraud, crimes_against_children, criminal_forums, denial_of_service, employment, extortion, gambling, government_impersonation, hacktivist, harassment_threats_of_violence, investment, lottery_sweepstakes_inheritance, malware_scareware_virus, misrepresentation, non_payment_non_delivery, overpayment, personal_data_breach	Non-platform-based	Tools
	criminal_forums	Platform-based	
Subject loss	[2]Ipr_copyright_and_counterfeit, lottery_sweepstakes_inheritance, advanced_fee, auction, bec_eac, confidence_fraud_romance,	Non-Exploitative	Attacker (Perpetrators)
	corporate_data_breach, credit_card_fraud, denial_of_service, employment, extortion, government_impersonation, harassment_threats_of_violence, health_care_related, identity_theft, investment, ipr_copyright_and_counterfeit, lottery_sweepstakes_inheritance, non_payment_non_delivery, other, overpayment, personal_data_breach, phishing_vishing_smishing_pharming, ransomware, re_shipping, real_estate_rental, social_media, tech_support, virtual_currency		Objective

[2] See meaning of abbreviations at https://www.ic3.gov/about/default.aspx

	crimes_against_children, malware_scareware_virus	Exploitative	
Subject count	Personal_data_breach, corporate_data_breach, civil_matter, harassment_threats_of_violence, misrepresentation, crimes_against_children, identity_theft, other, social_media, auction, real_estate_rental, health_care_related, credit_card_fraud, investment, ipr_copyright_and_counterfeit, malware_scareware_virus, virtual_currency, confidence_fraud_romance, ransomware, non_payment_non_delivery, terrorism, tech_support, denial_of_service, employment, overpayment, lottery_sweepstakes_inheritance, advanced_fee, phishing_vishing_smishing_pharming, bec_eac, charity, re_shipping, criminal_forums	Non-nation State	Attacker (Perpetrator)
	Government_impersonation, hacktivist, extortion, gambling	Nation-State	