International Innovative Research Journal of Engineering and Technology
ISSN: 2456-1983 Vol: 5 Issue: 3 March 2020

Designing Bots for Investigating Online Finanacial Frauds

Madiajagan M[1], Dheeba J[2]

[1,2]Associate Professor, School of Computer Science and Engineering,
Vellore Institute of Technology, Vellore

Abstract: Online financial frauds are one of the leading issues in the fields of digital forensics and cyber security today. Various online firms have been employing several methodologies for the prevention of finance-related malpractices. This domain of criminal activity is becoming increasingly common in the present cyberspace. In this paper, we will try to implement online financial fraud investigation using the digital forensics tool: Autopsy. A few existing cyber- security techniques for the investigation of such crimes, namely the Formal Concept Analysis and Confirmatory Factor Analysis; have been analyzed and reviewed. These techniques are primarily based on mathematical cyber security concepts. Henceforth, it has been tried to find out whether the investigation of similar crimes can be done satisfactorily using the readily-accessible digital forensics tool: Autopsy. Also, it has been explored whether the aforementioned cyber security techniques can be embedded along with the digital forensics tool to achieve the best results, by means of training a bot to automatically perform accurate investigations of such crimes. Thus, it has been tried to automate the process of online financial fraud investigation.

Keywords: Cyberspace, Computer crime, Forensics, Computer security, Software tools

1. INTRODUCTION

Online frauds refer to the usage of Internet services or other open-source software requiring Internet access to frame users or to otherwise take advantage of them. Finance-related flaws are becoming quite commonplace today. Digital forensics is a subsidiary of the main discipline “Forensic Science”. While forensic science includes all studies, techniques and findings related to various types of crimes in all domains; the field of digital forensics is solely concerned with the investigations involved in the sphere of cyber-crime. It is used for the recovery and analysis of the various computational devices suspected to be involved in the crime; or found at the crime scene. Cyber security is a discipline which is primarily aimed at the overall protection of computer systems within an organization. This includes the security of the system software, hardware as well as data stored in the system database. Thus, it basically protects computational assets from online and cyber-attacks. There are various techniques employed for the fulfilment of the purpose of safety of the cyberspace, such as computer access control, insertion of safety codes, compulsory authentication, encryption, firewall, etc.

The Formal Concept Analysis is a mathematical cyber security technique. It works by first structuring the input dataset into a concept lattice; and then the division of these formed lattices into a binary lattice. This binary lattice can be then used to verify which data is fake; and which one is presumably fraud. This technique has already been applied to the field of online financial fraud investigations, with satisfactorily good performances.

In the Confirmatory Factor Analysis, the primarily quantitative-type data analyzed by the investigators is then checked with the previously-existing similar data records; in order to “confirm” the consistency of the result interpretation. In case the deviations are very small, this technique is extremely useful to predict the forensic results. This technique has also already been used in the field of online financial crime investigation.

Autopsy is a fairly popular digital forensics tool. It serves as a platform as well as graphical interface for digital-crime investigation purposes. It is a much simpler part of the Sleuth Kit forensics software. It is primarily used by corporate organizations, law enforcement agencies and to some extent, by the military forces – for online crime inspection. It can be used to extract the past events that occurred on a particular computer system. It provides features such as creation of disk images to prevent evidence loss, analysis of user activity, analysis of the discovered data on the system, retrieval of the deleted data items, etc. It is supported by most operating system platforms; such as Windows, Ubuntu, Linux, Unix, etc. It can also be run on the cell phone platform of Android;

using the specialized "Autopsy: Mobile Forensics" toolkit.

Today, artificial intelligence is being extensively used to train robots to perform various kinds of functions with satisfactory accuracy. The AI bots can be trained using the "supervised learning" technique, in which the past cases and their results are shown to the bot. Then, the bot uses this "pattern" to predict the result in case of a similar new case. Since a bot would function faster and more efficiently than a human, this would make the investigation of cyber crimes and utilization of digital forensics tools faster and more accurate.

The most common types of online financial frauds include:

Phishing: Here, the fraudsters acquire user's sensitive data such as passwords and credit card credentials by means of email messages, fraud websites and phone calls.

Card Skimming: This crime involves the illegal extraction of the user's sensitive financial details on the magnetic stripe from ATMs, debit and credit card. This is usually done by installation of malware on the card reader used by the victim.

Smishing: It involves the extracting of a user's bank account details by means of exchange of text messages over the cell phone.

Vishing: It is the theft of sensitive data using information interchange over the phone; either by messaging, instant messaging or phone-calls, etc.

SIM Swap fraud: Here, the attackers replace the victim's SIM card with a false one; or are able to somehow install spyware on the original SIM card – which enables them to access all of the user's phone data, including those related to finances.

Identity theft: In this crime, the criminal, using some basic stolen credentials of the victim, such as Date of Birth, phone number, addresses and credit card numbers, builds up a fake identity posing as the original victim.

Internet-based financial crimes steal millions of dollars each year from the victims, and continue to terrorize the Internet through various approaches.

2. LITERATURE SURVEY

Recently, various authors has proposed algorithms for online fraud detection. Antoine Bouveret (2018), discussed the growth of online crimes related to the field of finances. It was found that the number of such crimes as well as the emergence of "new crime variations" were growing rapidly. This demonstrated the urgent need for reliable cyber security and digital forensics techniques to curb this internet threat.

Tommie W. Singleton (2006), discussed about the significance of cyber evidence in various criminal activities, including financial malpractices. He concluded that digital forensics is an increasingly important area in the investigation of several crimes using forensic sciences. According to him, digital evidence should not be neglected as it can give new directions to any legal proceedings. Also, the digital investigation should not be limited to just the victim and accused's computational devices; but also, be extended to other peripherals.

Matthew Kul and Nick Waler (2017) discussed about the growing importance of cybersecurity techniques and tools, with respect to the rapid spurt of growth of cyber frauds, especially financial frauds. They also discussed the various challenges involved the domain of cybersecurity, such as legal permissions, unresolved technical issues, etc.

At the Digital Forensics Research Workshop (DRFWS)'s conference proceedings held in 2006 in the USA, Dr.Simson Garfinkel proposed the technique of Cross Drive Analysis for digital crime investigation. In this method, data could be accumulated from various suspect drive's (or other sources). The accumulated data was then statistically analyzed and correlation between them was found out. This trained the model to correlate any input data with a particular pre-defined category.

Benjamin E. Onodi, et al (2015), explored Garfinkel's technique of Cross Drive Analysis for investigation of financial cyber-crimes. For the implementation of this experimentation, data comprising of credit card numbers, email addresses, and other kinds of confidential information was accumulated from various victimized hard drives and other sources; and their correlation with the perpetrator's communication messages, geographical coordinates, etc. was found out. This gave a significantly clear idea about the actual perpetrator.

Frano Škopljanac-Mačina, et al (2013), discussed this mathematical investigative procedure in detail. In this paper, the working of the technique and its basic principle was briefed. According to the paper, the technique generates "concept lattices" based on the input datasets, in which similar inputs are grouped into one lattice. Thus, the input data can be divided or classified under various label names, which makes it a useful tool for forensic sciences as well.

Waziri et al (2014), discussed the application of the mathematical cyber security-based technique of using Formal Concept Analysis (FCA) for the binary classification of the input dataset into either genuine or fraud. In this model, The FCA technique was used to analyze the various data gathered from victim as well as suspects' mobile communication devices such as cell

phone, tablets etc. Then, the visualization of the relationship between the crime occurrences within different proximal geographical areas was achieved successfully. This helped to develop a pre-trained model which, when given similar crime-investigation input as well as geographical area, could classify the data as fraud or not. This would greatly help financial firm websites.

Peter Prudon (2015), provided a detailed explanation and criticism related to the usage of the Confirmatory Factor Analysis technique in investigation procedures. He has discussed about how the methodology of calculating deviation between the predicted results and the previously-known results (relating to similar research cases) could be used to determine the accuracy of a particular prediction. Such a prediction can be used in investigative sciences. Also, the extent of accuracy of results was discussed; which was satisfactory.

Hamdan MW (2018), applied the Confirmatory Factor Analysis for the investigation of financial e-frauds. For this implementation, firstly, previously collected data related to completed cases of online financial frauds was collected and analyzed. Then, for new such cases based on somewhat similar patterns, results related to the impact, perpetrators, future possibility of attacks, etc. were predicted manually. Then, whether the prediction is accurate or not was determined using the deviation calculated by the Confirmatory Factor Analysis. If the deviation obtained was acceptably small; the conclusion was that correct investigative results had been achieved.

Nisarg Trivedi and Dhruv Patel (2015), discussed the various features provided by the forensic tool: Autopsy. They analyzed the software's efficiency using various test cases. They have also described the functioning of the software for the cases they investigated using it. They concluded that Autopsy was fairly well-performing when it came to conduction of digital investigations; with limited number of issues.

Simson Garfinkel (2010), discussed about the features provided by the forensic tool: EnCase. He discussed its working, versions, features, limitations, etc. On the whole, he concluded that, as of now, EnCase is one of the topmost available forensics tool, which is heavily reliable and easily accessible for various types of cyber-crime cases.

Adam Cervellone, et al (2019), compared various open-source digital forensics tools. According to the results obtained by the paper, it was found that EnCase was one of the leading and most reliable forensic tools. On the other hand, Autopsy was also quite a well-performing software for cyber-crime investigation.

M.P. Wellman (1995) in the survey paper "The economic approach to artificial intelligence" concluded that the application of artificial intelligence, for various practical uses, could be done in such manner so as to keep the required expenditure as low as possible. This can be used for the design of cheaper AI bots.

G. Lakemeyer & B. Nebel (2003), gave a brief summary about the various artificial intelligence techniques such as those related to clustering, classification, decision-making, etc. These techniques can be applied to the various bots embedded with AI.

T. C King, et al (2019), proposed the concept of use of artificial intelligence to train devices to detect the suspects and criminals using the various algorithms which are already available for supervised learning.

3. PROPOSED WORK

Most finance-related online crimes are committed by means of first provoking the user to somehow give out their credentials, such as credit card number, password, etc. Thus, the crime can be detected in two major steps:

Extraction of emails and messages found to be provoking the victim to give out his credentials: This can be done on the victim's computer or phone. With the help of this step, the fraud email-id or phone number can be identified, which will ultimately help to track down the location of the criminal.

Once the crime suspects are identified, their computer's hard drive must be scanned for the victim's credit card information as well as evidence of sending messages to the victim in the first place.

Here is how Autopsy can be used to for such crime investigation:

Step-1: Extraction of suspect emails from the victim's computer drive depicted in fig 1.1.

Step-2: Extraction of evidence (victim's credentials) from the suspect's computer drive.

AI bots can be trained to utilize Autopsy for email and information retrieval, and then classify the emails as suspicious or not (Formal Concept Analysis). Then, the bots may also display the accuracy of their prediction, based on the source used for email extraction (Confirmatory Factor Analysis). This functioning is depicted as below:

Initially, the bot will create a disk image file of the victim's computer hard drive; and feed it to the Autopsy software as the input source file.

Then, using the steps mentioned before, the bot will retrieve all the emails which contain the suspected keywords (the emails which provoked the victim to disclose his bank credentials).

These emails are retrieved primarily either from the location "Outlook.pst" or "Windows mail". Based on

previous similar cases, the bot has been trained (through supervised learning), to classify which emails are more suspicious and important (based on the number and type of keywords, and the location from where they are retrieved). This binary classification is a cyber security technique called as Formal Concept Analysis.

The most suspicious emails are then used to track the sender's IP address. Subsequently, whenever the criminal connects to the internet, his location will be disclosed. Even if he uses a VPN (Virtual Private Network), the ISP (Internet Service Provider) can retrieve which VPN is being used, and the VPN company can disclose the suspect's location.

Then, after the suspect's computer drive is seized, the bot will retrieve all the deleted files, and search the entire system for information related to the victim (such as personal credentials, etc.), and to the crime (such as emails relating to theft, etc.). This is done using the Autopsy software.

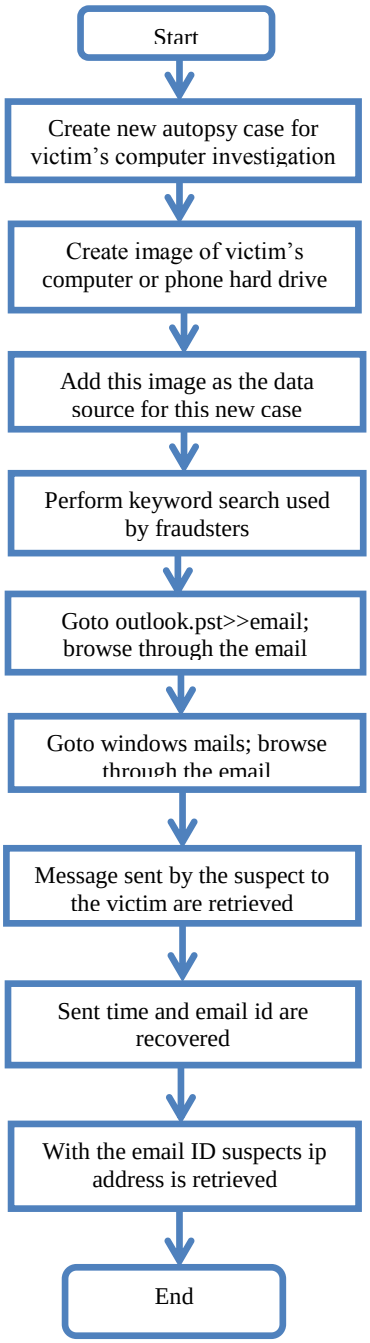

Figure 1. Procedure to obtain suspicious emails from the victim's hard drive

The bot then carries out a similarity test between the information retrieved from the victim as well as suspect's drives. This helps to predict the accuracy of the prediction (as to whether the suspect is the actual criminal or not). The accuracy calculation is done by the bot using the cyber security technique called as Confirmatory Factor Analysis, in which accuracy is predicted based on past similar cases.

4. CONCLUSION

We can conclude that to handle such a large number of finance-related cyber-crimes, AI bots can be trained to predict who has committed the crime. This can be done by embedding the forensics software "Autopsy" within the bot's processor; as well as training the bot (via supervised learning) to classify the emails and predict the accuracy of the results obtained using cyber security techniques (Formal Concept Analysis and Confirmatory Factor Analysis).

This mechanism will have several advantages. Usage of the digital forensics tool alone does not guarantee the accuracy of the results; and usage of the cyber security technique alone is a lengthy process. Moreover, the use of bots would save a lot of time and manpower.

The disadvantage of the proposed bot design technique is that it is highly resource-intensive. Development of AI bots alone requires a lot of technical resources. Providing them further training would incur even greater costs. Plus, as the system is new – it is more prone to glitches, which will be eventually resolved over time. Thus, if such a bot is successfully designed to investigate the online financial frauds, it would be greatly helpful to the investigate agencies.

REFERENCES

[1] Bouveret A. Cyber risk for the financial sector: a framework for quantitative assessment. International Monetary Fund; 2018 Jun 22.

[2] Michel P. Financial crimes: the constant challenge of seeking effective prevention solutions. Journal of Financial Crime. 2008 Oct 10;15(4):383-97.

[3] Kuhlmann S, Merkel R, Dittmann J, Zitturi BC, Griesbacher M. Criminals cash flow strategies in financial crime on the example of online and offline fraud. InThe European Conference on Psychology & the Behavioral Sciences 2016 2016 (pp. 61-71). The International Academic Forum (IAFOR).

[4] Bilge L, Strufe T, Balzarotti D, Kirda E. All your contacts are belong to us: automated identity theft attacks on social networks. InProceedings of the 18th international conference on World wide web 2009 Apr 20 (pp. 551-560). ACM.

[5] Onodi BE, Okafor TG, Onyali CI. The impact of forensic investigative methods on corporate fraud deterrence in banks in Nigeria. European Journal of Accounting Auditing and Finance Research. 2015 Apr;3(4):69-85.

[6] Škopljanac-Mačina F, Blašković B. Formal concept analysis–overview and applications. Procedia Engineering. 2014 Jan 1;69:1258-67.

[7] Waziri V, Olalere M, Umar A. E-fraud forensics investigation techniques with formal concept analysis. International Journal of Cyber-Security and Digital Forensics. 2014 Oct 1;3(4):235-45.

[8] Prudon P. Confirmatory factor analysis as a tool in research using questionnaires: A critique. Comprehensive Psychology. 2015 Jan 1;4:03-CP.

[9] Trivedi N, Patel D. Digital Evidence Handling Using Autopsy. Int. J. Sci. Adv. Res. Technol. 2015;1(1):10-8.

[10] Garfinkel SL. Digital forensics research: The next 10 years. digital investigation. 2010 Aug 1;7:S64-73.

[11] King TC, Aggarwal N, Taddeo M, Floridi L. Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. Science and engineering ethics. 2019 Feb 14:1-32.

[12] Alaieri F, Vellino A. Ethical decision making in robots: Autonomy, trust and responsibility. In International conference on social robotics 2016 Nov 1 (pp. 159-168). Springer, Cham.

[13] Russell SJ, Norvig P. Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,; 2016.

[14] Lakemeyer G, Nebel B, editors. Exploring artificial intelligence in the new millennium. Morgan Kaufmann; 2003.

[15] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015 May;521(7553):452-9.

[16] Wellman MP. The economic approach to artificial intelligence. ACM Computing Surveys. 1995 Sep 1;27(3):360-2.

BIOGRAPHIES

Prof. Dr. M. Madiajagan completed his B.E. Computer Science and Engineering from Madras University, Chennai, M.S. in Software Systems from BITS Pilani, Pilani, Rajasthan. He completed his Ph. D from BITS Pilani, Pilani, Rajasthan in the year 2009. Presently he **is** working as Associate professor in School of computer Science and Engineering, VIT, Vellore having rich teaching and research experience of more than 20 years and also served for 4 Years as Head of Technology in an International IT Industry and consultant many International IT Projects. He published papers in many National and International journals, conferences and book chapters indexed by Elsevier, Scopus, Springer, dblp, SCIE and many more. He areas of interests are Distributed Systems, Cyber Physical Systems, Brain Computer Interfaces, Network Security, Medical Imaging, Deep Learning.

Prof. Dr. J. Dheeba, completed her B.E. and M.E. in computer Science from Anna University Chennai. She completed her Ph.D from Anna university Chennai in the year 2013. Presently she is working as Associate professor in School of computer science and engineering, VIT, Vellore having rich teaching and research experience of 12+ years. She published papers in many National and International journals, conferences and book chapters indexed by Scopus, Springer, dblp, SCIE and many more. Her areas of interests are Algorithms, Medical Imaging, Deep Learning.