International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Data Behavioral Pattern Analysis and Noise Classification in Mobile Forensic

Preeti Dudhe[1], Dr. S. R. Gupta[2]

Submitted: 08/09/2023 **Revised:** 21/10/2023 **Accepted:** 06/11/2023

Abstract: Using data gathered from mobile devices and suspicious pattern detection algorithms, Mobile Forensic Data Analysis seeks to identify criminals. When criminal activity is associated with robotic processes like malware distribution, it is easy to foresee. When human behavior is involved, as in traditional crimes, prediction and detection become more alluring. Cyberbullying and small-scale drug sales, both of which rely heavily on mobile devices, are the focus of the current study, which proposes a combined criminal profiling and suspicious pattern identification methodology. Our evidence-based method improves accuracy and decreases false positives on a dataset by linking several observed patterns. After evaluating and displaying the results of tests conducted on a dataset that includes both benign and malicious traffic, the scenarios are re-run on a real dataset for further testing and verification.

Keyword: *Data analysis of forensic evidence, criminal profiling, analysis of behavioral evidence.*

1. Introduction

One of the youngest and least established areas of Mobile Forensics (MF) is mobile forensic data analysis. "Fallen into the trap of focusing almost exclusively on the collection of data and has paid very little attention to the examination and analysis phases," as M. K. Rogers puts it. Numerous scholarly works have pointed out the importance of adopting smart solutions that eliminate the need for time-consuming manual investigations [6]. Moreover, a questionnaire-based survey by Al Fahdi et al. [1] concerning the contemporary issues faced by forensic practitioners suggests that 85% of the participants annotated the "need to develop approaches to identify and extract significant data through techniques such as criminal profiling" [1] as an important issue. There is a dual function for forensic analysis of data and metadata. In the first place, it serves as a backup plan in case direct access to evidence fails to yield a tangible result owing to anti-forensic scenarios like data cascading or willful data tampering like criminals exchanging encoded verbal communication. Second, and more pertinent to this study, it can be used to relieve investigators of some of their duties by acting as a triage mechanism for possibly suspicious user behavioral patterns before or after a manual inquiry.

In contrast to previous research, we aim to provide a framework that can gather as much evidence as possible for a wide range of attacks and malevolent behaviors. This paper delves into the challenge of evidence collection by analyzing the structure of mobile forensic data. Indicators of criminal activity are the most prominent current developments. By keeping an eye out for these patterns, anomalies that are used by various attackers can be isolated. We use a pragmatic approach to discovering anomalous patterns by analyzing Mobile Forensic Data patterns with regression models to identify indicators of wrongdoing. The efficiency of monitoring in relation to attacks and other harmful behavior may be enhanced through correlation and analysis of the aberrant patterns that form evidence.

The remaining sections of the paper are structured as follows. In Section II, we conduct a literature review; in Section III, we discuss the approach employed in this research, providing a detailed, in-depth explanation of each step. The experimental findings are presented in Section IV. Section V finishes the report by discussing the most important findings and areas for growth.

[1]*Assistant Professor*
Department of Information technology
Prof Ram Meghe Institute of technology & Research, badnera
[2]*Assistant Professor*
Department of Computer science and engineering
Prof Ram Meghe Institute of technology & Research, badnera

2. Related work

Behavioral Evidence Analysis (BEA) approaches and automated criminal profiling are used in a large number of published studies that make use of intelligent computer methodologies. Most of them employ qualitative research and demographic information to draw conclusions. In addition, they are primarily interested in either contributing to or verifying the accuracy of an existing body of information. The NNPCP project [62] was one of the earliest efforts in this field; it involved the development of a NN that could profile criminal behavior across a variety of offences. The Italian police department's official database of criminal records served as its primary source of information. Despite the very remarkable description for the time the paper was written, the description of the NN architecture and functions is rather abstract and no more details on obtained results are offered. "Model criminal behavior from post-mortem databases of single-victim homicides" [18] was the goal of Ferrari et al.'s usage of Bayesian and Feed-Forward NNs, with the results of each approach being compared. The algorithms used as inputs a wide variety of psychological elements related to the offender's character and the specifics of each crime's execution, and then mapped those categories onto distinct outcomes. In order to construct a sociodemographic and behavior-based cybercriminal profile from an input set of crime types and their associated actions, Enache et al. [17] developed a multilayer NN. The authors state that some effective connections were made; nevertheless, they do not provide comprehensive results. The study by Islam and Verma employed Fuzzy Logic ideas to evaluate the safety of communications sent and received in a 3G network between different entities based on their identities and motivations. The system's inputs were the various permutations of the sender's familiarity with the device's owner and the device in use. The output of the system was the cumulative risk associated with each possible set of inputs. The danger level per SMS ranged from 0 to 5, with 0 being the lowest and 5 the greatest. Lai et al. presented "a conceptual framework for profiling internet pirates" that takes into account the individuals' technological habits and personalities. Based on "the facts," "the behavioral characteristics," and "the personality particulars," the writers crafted a profile of the internet pirate. Inferred from the available data, the facts included things like timestamps and the sorts of files that were traded. Personality particulars included more nebulous concepts including personality features, motivations that lead to piracy, and influences that created the potential pirate's profile, whereas behavioral characteristics focused on how a pirate behaves while using the Internet. Then, they developed and sent out a questionnaire with questions about the three topics described earlier in this paragraph. They utilized an approach called multidimensional scaling (MDS) [10] to generate pirate profiles by grouping files together based on shared attributes. An innovative feature of the new antimalware solution Andro-AutoPsy. In order to build criminal profiles, the programmed utilizes "similarity matching in malware creator-centric information" in addition to data regarding the technical aspects of the malware in question. Smali opcodes, AndroidManifest.xml metadata, and the serial numbers of various certificates are common places to find clues about the creator's intent. Andro-Autopsy is a hybrid detection engine that combines a rule-based, behavioral detection module with a classification engine that compares a sample's behavior to that of known malware. Quick and Choo presented the Digital Forensics Analysis Cycle (DFIAC), a comprehensive process paradigm for smart MF. Information from mobile devices confiscated by the South Australian Police between the years 2000 and 2015 was retrieved using the model. The authors effectively established connections between several criminal organizations. The development of criminal profiling studies over the years is readily apparent. More thought and substance have gone into more recent works. They also include a greater quantity of tests and outcomes, which serves as more convincing evidence. Finally, the writers employ a variety of approaches and draw on multidisciplinary concepts to strengthen their work. Their inability to do both profiling and detection is a major drawback. Andro Autopsy is the only real exception here, but even it has a stronger emphasis on machine behavior. The approach taken in this research is a synthesis of previous profiling methods and a brand-new suspicious pattern identification method built on NNs and ANFIS. The next part provides an in-depth explanation of the chosen methodology.

3. Our proposed system has the following benefits:

A system that represents static data will not be flexible enough to deal with the ever-changing nature of Mobile Forensic Data. First, it is impossible to teach typical behavior using neutral examples. It is very challenging to acquire a big, up-to-date, realistic Mobile dataset that exclusively includes benign flows and reflects the dynamic aspects of benign pattern. Our approach, on the other hand, is based entirely on information obtained from malicious patterns. Depending on the nature of the attack, the typical pattern of mobile forensic data will look very different.

Second, Mobile Forensic data analysis is successfully gathering proof of the attack thanks to our technology's ability to detect various abnormal patterns. For a variety of reasons, a single attack may exhibit multiple patterns.

We evaluate our approach using data collected from a variety of sources, both legitimate and malicious.

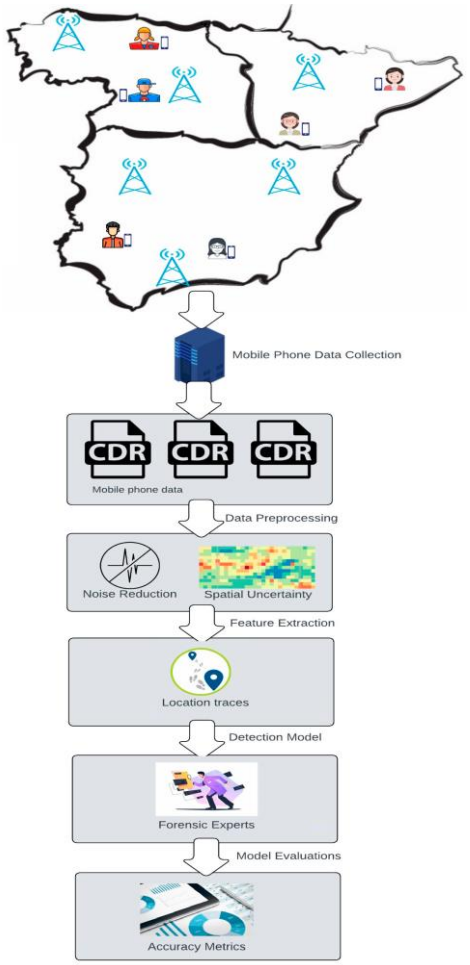

Fig. 01: working model

Our well-constructed evidence-gathering technique allows for a high detection rate of malicious traffic with a low false positive rate. Our analysis, in contrast to prior attempts, details the amount of evidence gathered for malicious Mobile Forensic Data sessions, which is useful for operators when deciding on a set-up.

Cell tower location data provides a wider sample of the general population and can show the positions of victims, perpetrators, and visitors at the time of a crime was committed, while data from individual mobile phones can represent the movement and communication records of suspects. This is the initial phase, and it begins with this. Researchers are able to investigate the links between human dynamics and interactions and spatial and temporal crime patterns with the help of crime scene data, which provides spatiotemporal information on crime episodes based on official records. The second stage, called preprocessing, entails the extraction of stay sites for home location recognition and other significant location and spatial mapping techniques for intersecting or projecting mobile network cells into spatial units. In the third stage, feature extraction, spatiotemporal features and call features are extracted and used to characterize criminal behaviors.

This second stage consists of two sub-steps: analysis and validation. Using various machine learning classifiers, analysts construct a detection model to identify illegal actions based on spatiotemporal and phone use variables. The best classification algorithms are chosen because they are the ones most likely to succeed. The data is then used in conjunction with social network analysis metrics and technologies to create a criminal ring of prospective bad actors. Data generated by the detection model is then analyzed by forensic experts to see how accurate the model is.

4. Results

Our well-constructed evidence-gathering technique allows for a high detection rate of malicious traffic with a low false positive rate. Our analysis, in contrast to prior attempts, details the amount of evidence gathered for malicious Mobile Forensic Data sessions, which is useful for operators when deciding on a set-up.

Cell tower location data provides a wider sample of the general population and can show the positions of victims, perpetrators, and visitors at the time of a crime was committed, while data from individual mobile phones can represent the movement and communication records of suspects. This is the initial phase, and it begins with this. Researchers are able to investigate the links between human dynamics and interactions and spatial and temporal crime patterns with the help of crime scene data, which provides spatiotemporal information on crime episodes based on official records. The second stage, called preprocessing, entails the extraction of stay sites for home location recognition and other significant location and spatial mapping techniques for intersecting or projecting mobile network cells into spatial units. In the third stage, feature extraction, spatiotemporal features and call features are extracted and used to characterize criminal behaviors.

This second stage consists of two sub-steps: analysis and validation. Using various machine learning classifiers, analysts construct a detection model to identify illegal actions based on spatiotemporal and phone use variables. The best classification algorithms are chosen because they are the ones most likely to succeed. The data is then used in conjunction with social network analysis metrics and technologies to create a criminal ring of prospective bad actors. Data generated by the detection model is then analyzed by forensic experts to see how accurate the model is.

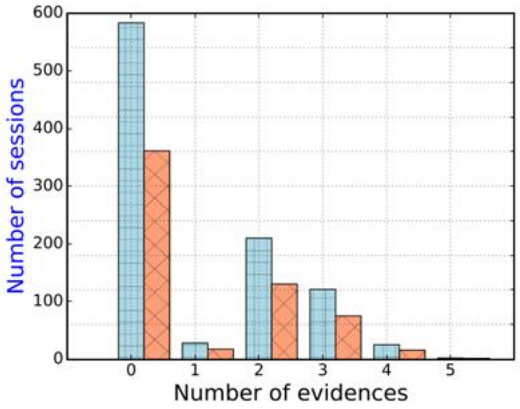

Fig. 2. Evidence histogram for typical mobile data sharing with specific periods.

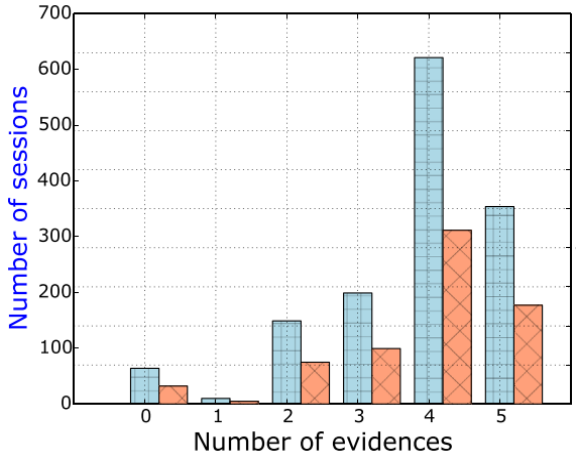

Fig.3: Distribution of malware mobile data sharing evidence histogram.

5. Conclusion

In this study, we introduce a system that uses pattern detection to gather indicators of attack and other bad behavior in mobile data sharing patterns. In our approach, typical pattern features are not required. Instead, we relied on regression models to investigate potential outliers. Experiments used a variety of data-sharing patterns, from "normal" sessions to those infested with malware. Our method allowed us to detect malware in Mobile forensic data sharing while reducing the number of false positives.

We hope to refine our method so that it can pick up clues about subsequent assault stages, such as infiltration and (successful) exploitation, in addition to the initial reconnaissance stage. Attacks can be detected, identified, and profiled with the use of mobile forensic data patterns with such a solution. In the future, we hope to try out more advanced features and correlation data, such as the closeness of mobile IMEI numbers, malicious domains visited, etc.

The real-time processing of live data sharing patterns is another goal of our framework development. There are a lot of hurdles that need to be cleared before that can happen. Fast analysis algorithms are required due to the large session size and large number of data sharing sessions.

References:

[1] Okmi, M.; Por, L.Y.; Ang, T.F.; Ku, C.S. Mobile Phone Data: A Survey of Techniques, Features, and Applications. Sensors 2023, 23, 908.

[2] Lanza, G.; Pucci, P.; Carboni, L.; Vendemmia, B. Impacts of the Covid-19 pandemic in inner areas. Remote work and near-home tourism through mobile phone data in Piacenza Apennine. TEMA 2022, 2, 73–89.

[3] Zhang, A.; Bradford, B.; Morgan, R.M.; Nakhaeizadeh, S. Investigating the uses of mobile phone evidence in China criminal proceedings. Sci. Justice 2022, 62, 385–398.

[4] Wu, J.; Abrar, S.M.; Awasthi, N.; Frias-Martinez, E.; Frias-Martinez, V. Enhancing short-term crime prediction with human mobility flows and deep learning architectures. EPJ Data Sci. 2022, 11, 53.

[5] Xu, Y.; Li, J.; Xue, J.; Park, S.; Li, Q. Tourism geography through the lens of time use: A computational framework using fine-grained mobile phone data. Ann. Am. Assoc. Geogr. 2021, 111, 1420–1444.

[6] Rummens, A.; Snaphaan, T.; Van de Weghe, N.; Van den Poel, D.; Pauwels, L.J.; Hardyns, W. Do mobile phone data provide a better denominator in crime rates and improve spatiotemporal predictions of crime? ISPRS Int. J. Geo-Inf. 2021, 10, 369.

[7] Casino, F.; Dasaklis, T.K.; Spathoulas, G.P.; Anagnostopoulos, M.; Ghosal, A.; Borocz, I.; Solanas, A.; Conti, M.; Patsakis, C. Research Trends, Challenges, and Emerging Topics in Digital Forensics: A Review of Reviews. IEEE Access 2022, 10, 25464–25493.

[8] Rizvi, S.; Scanlon, M.; McGibney, J.; Sheppard, J. Application of Artificial Intelligence to Network Forensics: Survey, Challenges and Future Directions. IEEE Access 2022, 10, 110362–110384.

[9] Pawlick, J.; Chen, J.; Zhu, Q. iSTRICT: An Interdependent Strategic Trust Mechanism for the Cloud-Enabled Internet of Controlled Things. IEEE Trans. Inf. Forensics Secur. 2022, 14, 1654–1669. Available online: https://ieeexplore.ieee.org/document/8543871/ (accessed on 1 February 2023).

[10] Studiawan, H.; Salimi, R.N.; Ahmad, T. Forensic Analysis of Copy-Move Attack; Springer International Publishing: Cham, Switzerland, 2021.

[11] Javed, A.R.; Ahmed, W.; Alazab, M.; Jalil, Z.; Kifayat, K.; Gadekallu, T.R. A Comprehensive

Survey on Computer Forensics: State-of-the-Art, Tools, Techniques, Challenges, and Future Directions. IEEE Access 2022, 10, 11065–11089.

[12] Heartfield, R.; Loukas, G.; Bezemskij, A.; Panaousis, E. Self-Configurable Cyber-Physical Intrusion Detection for Smart Homes Using Reinforcement Learning. IEEE Trans. Inf. Forensics Secur. 2021, 16, 1720–1735.

[13] Ashraf, N.; Mehmood, D.; Obaidat, M.A.; Ahmed, G.; Akhunzada, A. Criminal Behavior Identification Using Machine Learning Techniques Social Media Forensics. Electronics 2022, 11, 3162.

[14] Bibi, M.; Abbasi, W.A.; Aziz, W.; Khalil, S.; Uddin, M.; Iwendi, C.; Gadekallu, T.R. A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis. Pattern Recognit. Lett. 2022, 158, 80–86.

[15] Ali, M.L.; Thakur, K.; Obaidat, M.A. A Hybrid Method for Keystroke Biometric User Identification. Electronics 2022, 11, 2782.

[16] Gundluru, N.; Rajput, D.S.; Lakshmanna, K.; Kaluri, R.; Shorfuzzaman, M.; Uddin, M.; Rahman Khan, M.A. Enhancement of detection of diabetic retinopathy using Harris hawks optimization with deep learning model. Comput. Intell. Neurosci. 2022, 2022, 8512469.

[17] Nadia, T.; Obaidat, M.A.; Rawashdeh, M.; Bsoul, A.K.; Al Zamil, M.G. A Novel Feature-Selection Method for Human Activity Recognition in Videos. Electronics 2022, 11, 732.

[18] Long, D.; Liu, L.; Xu, M.; Feng, J.; Chen, J.; He, L. Ambient population and surveillance cameras: The guardianship role in street robbers' crime location choice. Cities 2021, 115, 103223.

[19] Long, D.; Liu, L. Do Migrant and Native Robbers Target Different Places? ISPRS Int. J. Geo-Inf. 2021, 10, 771

[20] Hu, X.; Chen, H.; Liu, S.; Jiang, H.; Chu, G.; Li, R. BTG: A Bridge to Graph machine learning in telecommunications fraud detection. Future Gener. Comput. Syst. 2022, 137, 274–287.

[21] Xing, J.; Yu, M.; Wang, S.; Zhang, Y.; Ding, Y. Automated fraudulent phone call recognition through deep learning. Wirel. Commun. Mob. Comput. 2020, 2020, 8853468.

[22] Chu, G.; Wang, J.; Qi, Q.; Sun, H.; Tao, S.; Yang, H.; Liao, J.; Han, Z. Exploiting Spatial-Temporal Behavior Patterns for Fraud Detection in Telecom Networks. IEEE Trans. Dependable Secur. Comput. 2022, 1–13.

[23] Ana Rodriguez, Kristinsdóttir María, Pekka Koskinen Pieter van der Meer, Thomas Müller. Machine Learning Techniques for Multi-criteria Decision Making in Decision Science. Kuwait Journal of Machine Learning, 2(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/214

[24] Dasari, S. ., Reddy, A. R. M. ., & Reddy , B. E. . (2023). KC Two-Way Clustering Algorithms For Multi-Child Semantic Maps In Image Mining. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2s), 01–11. https://doi.org/10.17762/ijritcc.v11i2s.6023

[25] Raghavendra, S., Dhabliya, D., Mondal, D., Omarov, B., Sankaran, K.S., Dhablia, A., Chaudhury, S., Shabaz, M. Retracted: Development of intrusion detection system using machine learning for the analytics of Internet of Things enabled enterprises (2023) IET Communications, 17 (13), pp. 1619-1625.