Cyber Forensics: Cybercrime Investigation through Logs

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Haris Saeed

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor
Richard D. Jacques, Department of Engineering and Society

Introduction

With the age of the Internet and the ongoing technological advancements in society, trying to log every action taken by users has become nearly impossible. In the past, companies used logs for troubleshooting problems by seeing where their product went wrong, but now they are used for many functions such as performance checking, recording user actions, and checking for malicious activity (Singh 2010). Logs are a collection of entries, and each entry contains information related to a specific event that has taken place. An enormous amount of operational information is created each time a user interacts with software. Many companies use logs as they are easy to capture at a large scale, quick at doing so, and to better understand customers and provide personalized experiences (Wang 2016). Additionally, companies and cyber forensics use logs to search for malicious behavior and find user information through behavior analysis (Singh 2010). In this paper, I will research how beneficial these logs can be in cyber forensics and cyberattacks. More specifically, whether these logs can aid with crime investigations and improve the security of society by allowing analysts to uncover suspicious behavior quickly to resolve and prevent crime.

Cyber Forensics

Cyber forensics is a new, rapidly changing field of technology that uses investigative techniques to extract information and data from a person's computer. It is a systematic approach for collecting and preserving data that guarantees information accuracy, reliability, and the presentation of the data in a legal environment. The goal of cyber forensics is to aid in investigating crimes. Evidence in the form of electronic information can be useful in solving cybercrimes (Lu 2017). This is applied to not only cybercrimes but also password breaking, spamming, data recovery and analysis, tracking user activity, forensic imaging and verification, viruses, file types, encryption, and more (Singh 2010). By applying cyber forensics to crime investigations, it can reduce the time needed to investigate and decrease the scope of the crime by reducing the complexity and narrowing the investigation. It is not meant to solve the crime itself.

Cyber forensics uses the same logs that everyday companies use to watch actions that users take on their platform and track their behaviors. Just like these companies, crime investigators can use behavioral logs to help with their investigations as stated before. Forensic analysis of log data is necessary to deal with cybercrime (Lu 2017). Companies have a goal of preventing and stopping hackers from accessing protected information and crime investigators have the job of hunting these hackers down. Many of these attacks such as bombings could be investigated by looking into purchases made by the bomber. On the internet in e-commerce, the client access server will generate data that has log files and query data (Wang 2016). Investigators and the seller can use the logs and see information like where the person is buying from and what they are buying, and function as needed. In a different scenario like a cyberattack, the firewall has logs that both the company and investigators can use to shut down the malicious attack and track down the attacker. It is not only companies that need logs to become much cleaner and filtered with only useful information being reported to them but also crime investigators.

Crime investigators and cyber security analysts can identify users and their behaviors from server log data for various cybercrime including phishing (Ibrahim 2021). Analysts take server logs and analyze them to determine whether an action is an irregular behavior and mark it down as malware as is done in network security applications (Chen 2005). One way they can determine whether an action is irregular is by looking at patterns that show various user intents.

The user’s intent is what the user’s goal is when using a program. Intents can differ between users and programs can be used for many different intents. When a person is using software with an intent, their intent strongly influences their behavior as they are more focused and engaged in a specific task and search through specific categories as compared to other users (Cheng 2017). By analyzing user actions with the timestamp, their intent can be discovered. Through intents, a user’s gender, age, and categorical differences can be determined. Taking both the log information that contains timestamps, location through IP addresses, and more and using information gained through intent analysis by comparing everyday logs to suspicious actions, malicious action can be identified and stopped, and information about the suspect can be discovered. Researching behavioral logs and how it helps crime investigations can further support the technical research described in the previous section and support the need for better behavioral logging practices.

Cyber Forensics Procedure

There are four steps to the forensics process according to the National Institute of Standards and Technologies (NIST) standard. The first step is the collection of the data. In this step, all the information and data collection related to the event is identified, labeled, and recorded. The integrity of the data must also be validated to ensure they can use it in the next steps. The Chain of Custody is also started in this step. The Chain of Custody is the process of tracking the movement and control of an asset by documenting each party that was involved with the data, collecting the purpose of the interaction with the data and the timestamp of the interaction. The second step is the examination process. This stage involves forensic tools and techniques used to filter the data and categorize them into different groups. Relevant information is extracted from the collection step. Automatic and manual methods by experts are used in this step to do the examination. The results of this process are recorded and noted in the Chain of Custody, along with what evidence is being used and how it was used. The third step of the process is the analysis of the data filtered from the examination phase. The analysis is conducted using justifiable methods and techniques to find useful information to help resolve the investigation. The final step is the reporting of the findings. This step includes what actions were taken in the investigation, what tools and procedures were used, and what else should be done if there are still ongoing issues such as additional resources needed or different methods. It also includes any vulnerabilities that the investigation found in the data sources they analyzed and recommendations for fixing those problems (Kotsiuba 2019). A flow chart of this process can be seen below in Figure 1.

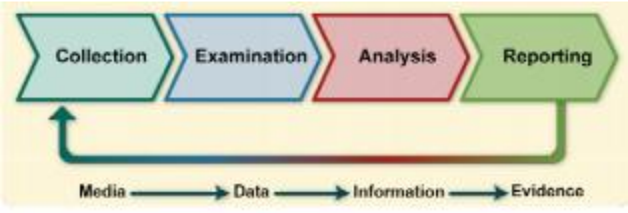

Figure 1. Cyber Forensics Process with Data Processing Sequence

To successfully conduct the cyber forensics process by analyzing logs, the system itself needs to have the infrastructure that allows the process to be conducted efficiently. During this process, there are several challenges for forensic analysts. One problem is the acquisition of the data. With so many sources of evidence within a network, there is a large amount of data that needs to be filtered to find useful data and evidence. Additionally, political, and technical limitations can also prevent access to data needed for the investigation. A technical limitation could be the limited storage in a device, which would result in missing data since the device can only capture a limited amount of information. Another issue is privacy and seizure of data. Due to the way network-based acquisition techniques are and the processing of personal data, certain privacy requirements are required to be followed when conducting the analysis. Finally, cyber forensics is a new investigative process that grows as our world becomes more technologically advanced. The legal system is still in the process of imposing laws for different types of procedures and techniques and there is a lack of precedent for cyber forensic analysts to follow, so the procedure must be done carefully with ensuring the integrity of the data and ensuring the privacy of all parties involved in the investigation (Kotsiuba 2019).

To follow the cyber forensics process and analyze logs efficiently while also dealing with the challenges in doing so, there are several good practices. The first is the identification and preservation of all data and any recent actions a party may have taken before the data could be collected. For example, emails, social media, and other data could be damaged and destroyed before analysts could access the data and preserve it. This information that was destroyed could be very important to the investigation, so identification of any actions taken by parties involved alongside the data collection and identification is important. The next good practice relates to the collection and processing of data. The data collected needs to ensure proper confidentiality of the party when revealing contents of the analysis to parties outside the investigation along with ensuring that the data is not forged and has integrity. When processing the data, a copy of the data should be used instead of the original data collected. This is to ensure the original data collected is untouched and can be reused at later steps or referred to. Additionally, when analyzing the data, many different tools should be used as there are a wide variety of tools available and different tools may yield different results or could help solve different objectives in the investigation. There are a few good practices when dealing with the review and reporting parts of the investigation as well. When reviewing the data that was used and the results from the analysis, analysts must ensure the data meet their criteria and objectives have been reached before sending a report to the parties involved. The production of the report is done by one party and then it is sent to related parties. The exact details of this depend on the investigation and any usage of the data, evidence, and results will be written out and agreed upon between the parties (Janloy 2020). Following these practices when conducting cyber forensic analysis of logs can ensure procedures are followed properly and challenges are minimalized as much as possible. The next section contains several case studies that follow these procedures and practices when conducting forensic analysis with logs to see if these logs that hold user behavior and actions can be useful with crime investigations and help improve the security of society to uncover suspicious behavior quickly to resolve and prevent crime.

Case Studies involving Cyber Forensics and Log Analysis

Control Risks Case Study: Network Data

The first case study involves internal company network data logs revealing the theft of company secrets. Xiaolang Zhang was an engineer with Apple's autonomous car division for 2.5 years, and he decided to resign and return to China to take care of his mother. He told the manager before leaving that he would be working with an electric car manufacturer in China. The manager was suspicious and ordered security to start an investigation. Through their investigation, they found nothing on his phones and laptop, but his network activity and data showed that there was a spike in activity in his final days working for Apple. They found that he downloaded many pages of information from secret databases that only company employees could access. The manager confronted him, and he admitted to taking company data and was indicted for theft of trade secrets (Control Risks 2021). In this case study with Apple, cyber forensics analysis with network logs showed how they found the suspicious activity and they were able to react swiftly and confront Zhang. Without these logs, Zhang would have gotten away with trade secrets and potentially helped the electric car company in China that he was going to go work for after leaving Apple. Forensic analysis of logs, specifically network logs, was helpful here in defending Apple's security and its trade secrets before the employee was able to leave for China and reveal those secrets.

Global Digital Forensics Case Studies: Intellectual Property and Drugs

The next two case studies involve a company called Global Digital Forensics (GDF), a cyber forensics company that helps companies by conducting cyber forensics for varying situations. The first case is an intellectual property case where two employees and a manager left a chemical manufacturing company. Before leaving, the employees and manager extracted financial data and proprietary information to their personal computers from the customer contact database. The employees believed that they were untraceable since they deleted their emails after collecting the data. The GDF helped the attorneys and identified several sources of evidence. They collected data from mail servers, backup tapes, the customer contact database, and more. The desktop systems were analyzed, and the original emails were recovered, allowing them to check dates and times to prove that the employees did steal information that they denied initially. The attorneys determined that they lost more than 150M USD, more than twice the amount they expected. In this case, GDF used cyber forensics and analyzed logs that they achieved from gathering data from various sources. They were able to use these logs to confirm their allegations and determine the actual amount the company lost because of the stolen information (Global Digital Forensics n.d.a). The use of cyber forensics and log analysis was not able to prevent damage from occurring, but instead, it was able to limit the amount of that that would have incurred if the employees had not been stopped, thus proving useful in uncovering suspicious behavior and resolving the crime investigation.

The second case study deals with a pharmaceutical company and its sales. A geographical region in Asia where normally a high volume of drugs was not selling as much anymore. They launched an investigation and found that the drugs were being diverted from those areas and instead being taken into the US and sold through other distributors, especially private nursing homes. The drug transportation was immediately halted along with the equipment involved. This process took some time due to the fact there was limited communication between the distribution companies. The FDA and local authorities conducted computer forensic analysis, but they found that the logs were severely backlogged, and the encrypted systems were not breakable. The GDF quickly responded and collaborated with the authorities to run their cyber forensics. They copied the hard drives from the warehouses where the drugs were transported to and from and made sure the authorities accepted their methodologies before conducting the forensic analysis. The GDF was able to decrypt and extra the information and extract data from the hard disks to show documentation on the transportation of the drugs. They found that there were distributors in Europe and Canada buying shipments and shipping to the US through repackaging and exporting. The drugs that were supposed to go to Asia were vitamins relabeled as drugs. This went on for 10 years before this investigation occurred. The company lost 13M USD each year this went on in counterfeit drugs throughout Asia and potentially endangering the health of their Asian community. The suspects were caught and put in jail in the US (Global Digital Forensics n.d.b).

This case study involved millions of lives across the globe and serious health issues. Asian drug consumers have been consuming vitamins for over 10 years when they needed the drugs for their treatments. Due to the mislabeling of drugs by illegal distributors, many lives were probably lost in Asia over time due to vitamins labeled as drugs that they were not. GDF was able to resolve the issues using varying methods of forensic analysis and extracting the information from distribution and warehouse hardware logs. Although they were able to stop the distributors and imprison them, the pharmaceutical company did not notice for more than 10 years resulting in the loss of money and human lives over that course of time. Cyber forensics and log analysis here were able to quickly determine the drug distribution trail and stop the illegal distribution from continuing, resolving the investigation. They were not able to stop it from initially occurring in the first place due to lack of knowledge of it happening until many years later.

Cyber Diligence: Unauthorized Emails and Network Forensics

The final two case studies involve a cyber forensics company known as Cyber Diligence. They offer the same services as the Global Digital Forensics company. The first case study involves unauthorized email access by the CIO that was found by the CSO of a defense contractor. The CSO claimed that the CIO of the company was unlawfully accessing the CEO and COO's extremely sensitive emails. To keep the CSO's identity anonymous while also not alerting the CIO that the investigation was occurring, they tried to stage a comprise call that the CSO usually gets and use this to check certain laptops for any data breaches. The plan worked and they forensically imaged twelve executives' laptops to avoid looking suspicious to the CIO. They found through the logs that the suspect was using a web-based interface to log into other executives' emails. The CIO was confronted and removed from his position (Cyber Diligence n.d.a). In this case study, forensic analysis was used to see whether the CEO was viewing sensitive emails that he should not have access to. They found through logs and their forensics that the CEO was, and he was invading the privacy of other executives and employees. It is not explained whether the CEO did anything with the information that he was accessing, but it was still an invasion of privacy. Through the analysis, the privacy and security of the employees were secured, and the company's security overall also improved through the termination of the CEO.

The second case study involves overseeing the network security of a government agency. Cyber Diligence was tasked to monitor a government agency's network security and so they applied a cyber forensics method on the internet gateway to analyze incoming and outgoing traffic. Every 24 hours, the data would be analyzed to see any suspicious activity. They noticed a workstation was accessing a blog usually used by hackers and at a closer look, found pornographic material and copyrighted movies illegally downloaded while there was little to no work being done. They told their client what was happening and got authorization to forensically image the computer for logs of history and activity and usage patterns. Analysis showed that what they saw has been happening over the past three months. The contracted employee was terminated, and the contractor reimbursed the client for all the fees that they charged for the year (Cyber Diligence n.d.b). In this case, log analysis as a form of cyber forensics was used to analyze network traffic and computer activity to find that a contracted employee was not doing work that they were paying him for. The company was able to regain a partial amount of the money that it had lost through paying that employee. Using cyber forensics, the company was able to prevent any further loss of money and secure the finances of the company, but they could not do more to regain their lost time and the remaining financial losses.

Conclusion: Efficiency of Cyber Logs and Society's Safety

This research paper attempts to answer whether logs can be beneficial in cyber forensics and cyberattacks and whether these logs can aid with crime investigations and improve the security of society by allowing analysts to uncover suspicious behavior quickly to resolve and prevent crime. Through case studies, we found that logs can be beneficial in cyber forensics and cyberattacks. Analyzing logs is one of the main methodologies used in cyber forensics. Although there are challenges, there are also good practices to combat those challenges to allow quick and easy log analysis to conduct cyber forensics. Cyber forensics and log analysis are crucial to solving cyber-related situations. By examining logs and analyzing them, losses and other dangerous consequences can be prevented. When conducting cyber forensics, certain guidelines must be followed that could be specific to that situation or a guideline that must be followed in any situation. To conduct the forensics in the first place, permission is needed from the client. Without the client's permission, the forensics are limited and not much can be done. Without permission, there is an invasion of privacy issue that could lead to legal consequences. Other guidelines might be the need for undercover actions as seen in Cyber Diligence's first case study where they had to avoid getting exposed by the CIO while also hiding the identity of the client.

Cyber forensics and log analysis can also help improve the security of our society by preventing crimes from escalating. A person must do something for a log to be created. There is no log if a person does nothing. This means that cyber forensics and log analysis cannot completely prevent crime, however, they can prevent the crime from becoming worse than it currently is. Additionally, logs take time to analyze which can delay crime investigations. In the Drug Diversion case study handled by Global Digital Forensics company, the logs were encrypted, and it delayed the investigation until they were able to decrypt the information. In other cases, cyber forensics, and log analysis were not conducted until a much later time from when the crime initially started, resulting in losses until the crime was exposed and dealt with. Overall, log analysis is useful in cyber forensics and helps prevent crime from escalating, but the crime must occur and be recognized as suspicious activity beforehand. They can aid with crime investigations and improve the security of society by enabling analysts to uncover suspicious behavior quickly and resolve crime investigations.

References

Cheng, J., Lo, C., & Leskovec, J. (2017). Predicting intent using activity logs. *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*. https://doi.org/10.1145/3041021.3054198

Ibrahim, K., Obaid, A. (2021). Fraud usage detection in internet users based on log data. *International Journal of Nonlinear Analysis and Applications*, 12(2), (pp. 2179-2188). doi: 10.22075/ijnaa.2021.5367

Lu, W. (2017) Exploration and implementation of user behavior forensics analysis system of computer network based on system log1. *2017 Institute of Thermomechanics*, 62(2), (pp. 53-62). Retrieved from http://journal.it.cas.cz/62(2017)-2B/Paper%2006%20Wenzhe%20Lu.pdf

Chen, S., et al. (2005), User Behavior Map: Visual Exploration for Cyber Security Session Data. *IEEE Symposium on Visualization for Cyber Security (VizSec)* (pp. 1-4). doi: 10.1109/VIZSEC.2018.8709223.

Singh, N. K., Tomar, D. S., & Roy, B. N. (2010). An Approach to Understand the End User Behavior through Log Analysis. *International Journal of Computer Applications*, 5(11), (pp. 27–34). https://doi.org/10.5120/953-1330

Wang, N., Zhang, Q., Yang, L., & Chen, M. (2016). A novel E-Commerce recommendation system model based on the pattern recognition and user behavior preference analysis. *Information Science and Industrial Applications*. Retrieved from https://web.archive.org/web/20180602064635id_/http://onlinepresent.org/proceedings/vol138_2016/23.pdf

I. Kotsiuba, I. Skarga-Bandurova, A. Giannakoulias and O. Bulda, "Basic Forensic Procedures for Cyber Crime Investigation in Smart Grid Networks," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 4255-4264, doi: 10.1109/BigData47090.2019.9006215.

K. Janloy and P. Boonyopakorn, (2020) "The Comparison of Web History Forensic Tools with ISO and NIST Standards," 2020 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Phuket, Thailand, 2020, pp. 1-4, doi: 10.1109/ITC-CSCC55581.2022.9894903.

Control Risks, (2021) “Five case studies with digital evidence in corporate investigations”. Control Risks. 2021, July 30, from https://www.controlrisks.com/campaigns/compliance-and-investigations/five-case-studies-of-interest-to-corporate-investigators

Global Digital Forensics. (n.d.a). Case study Worldwide Intellectual Property Case. Global Digital Forensics, from https://evestigate.com/case-study/worldwide-intellectual-property-case/

Global Digital Forensics. (n.d.b). Case study Drug Diversion. Global Digital Forensics, from https://evestigate.com/case-study/worldwide-intellectual-property-case/

Cyber Diligence (n.d.a) Case Study-5 Employee Misconduct Unauthorized Email Access By CIO. Cyber Diligence, from https://cyberdiligence.com/case-studies-5.html

Cyber Diligence (n.d.b) Case Study-6 Contractor Misconduct Network Forensics. Cyber Diligence, from https://cyberdiligence.com/case-studies-6.html