|| Volume 7 || Issue 8 || October 2023 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

FORENSIC INVESTIGATION AND PREVENTION ON VIRTUAL MACHINES

[1]Rutika Karande, [2]Tushar Mehetre, [3]Renuka Mehetre, [4]Nikita Ghadge, [5]Dr.Rokade M.D

[1-4]Sharadchandra Pawar College of Engineering Otur(Dumbarwadi),Tal - Junner, Dist - Pune . Pincode – 412409

[5]Sharadchandra Pawar College of Engineering Otur(Dumbarwadi),Tal - Junner, Dist - Pune . Pincode - 412409

Abstract

Cybercrimes is an illegal activity in which criminals use intelligent machines like computers and other network devices as their primary source for profiting by breaking the law. Cybercrime is a criminal offence. Cyber-attacks continue to increase, cyber-attacks detect and protective measures are often failing to track cyber-attacks via manual investigations. Machine learning thus plays a crucial role in the identification of cybercrimes. It has the capacity to track, evaluate, and avoid cyber-attacks to minimize the cyber-crimes incarnation. The application of machine learning methods such as clustering can thus help to develop an annual cybercrime detection system and cyber-attack prediction. There is a range of strategies in current cybercrime literature through feature extraction. In this context, a new system is proposed for cybercrime offenses by feature removals. Any unstructured cybercrime report can be uploaded to generate structure figures through a machine learning techniques in this proposed system. Subsequently the framework should include a report on the severity and occurrence of the categorization and resolution of cyber-crime offenses. The function summary is extracted using text mining algorithms and performance measurements and cybercrime prediction analyses

Introduction

Cloud storage is a modern concept that enables users to not only upload data to the web but also to have quick access to available resources and exchange data with others at any time. However, cloud is a technology that poses a challenge to those investigating and discovering forensic evidence that can assist in forensic investigation, as data stored on cloud can be accessed from anywhere and from any device, leaving very little traces. We cannot survive today without computers and the Internet because we rely on these devices for almost all of our jobs. All has been automated to computers, from home to education to banking and even corporate operations. All of our essential data is stored in digital format on computers. Because of their ability to simulate computing environments, isolate users, restore previous states, and facilitate remote initialization, virtual machines are rapidly gaining popularity. All of these features have a positive impact on defense. The VM's hardware abstraction and isolation limit the attack's reach and make it much more difficult for an attacker to gain unauthorized access to data and resources on the physical machine. Users may restore their virtual machines to a state prior to an attack or data loss, making malware removal and data preservation simple. By allowing users to start and stop virtual machines from afar, attackers have a limited amount of time to prepare and conduct their attack. This is an extremely effective security precaution. Hypervisors have the ability to search for malware because they run outside of the VM.

Literature Review

According to [1] looks at the possibilities of applying machine learning to identify harmful threats. In the field of cybersecurity, machine learning has a lot of promise. The random forest classifier surpasses all others on the dataset and gets excellent accuracy, which is a clear trend. The three best characteristics discovered are the length of the input, the amount of punctuation letters, and the number of different bytes. Malicious writes are more dangerous than malicious reads. As a result, the models were taught to distinguish between read and write operations. When attempting to identify malicious code, this work explores which attributes are excellent and which are harmful to employ in machine learning classifiers. This system has improved as a result of the discovery of new useful characteristics and the testing of bigger datasets.

According to [2] it is based on the study and surveys of bandwidth attacks that concentrate primarily on DDoS, which are truly a ruthless challenge and are difficult to detect and reduce the network's efficiency. In order to deter legitimate users, DDoS contains a community of attacker nodes and targets the victim. Through accessing the services and resources of the network. Procedures that are the mechanisms of intrusion prevention in IoT devices treated as Intrusion Detection System Add-ons to actively protect and avoid intrusions detected by the intrusion system. The IDS's identification procedures. The report that the IDS generates after evaluating the forensic investigation report is the cornerstone of the proposed method. This article highlights the possible safety strategy and suggests a possible safety technique a prevention mechanism that is useful for IoT networks that are prone to DDoS attacks. We have sued outcomes in the proposed algorithm, in a time-related manner based on the basic structure and functions of the current IDS.

According to [3] leads us to question the need for an advanced Digital Forensics Investigation System (DFIF) for the successful prosecution of digital crime in court; in such a way that, during the process, the framework can protect the credibility of evidence. Our paper is descriptive in nature, analysing recent cybercrime attack patterns and discussing relevant cyber forensics. Furthermore, we have mapped process and performance produced by various stages in the DFIF that have been examined from previously proposed frameworks and represented a comparative mapping of all framework frameworks. The mapping scheme provides a structured DFIF for the establishment of consistent forensic process/action guidelines and the acquisition of a precise performance idea for each unique activity that is associated during the investigation. In our analysis of the previously proposed structure, overlays of steps/processes with a different vocabulary, focus area and outline characteristics were defined at each point.

According to [4] a system for the study of chat logs with data mining and Natural Language Processing (NLP) techniques for crime investigation. The proposed structure collects chat logs from the social network and summarizes conversations into topics. In order to see the crime-related findings, the crime analyst may use the Information Visualizer. To assess the feasibility of our proposed system, we partnered with a Canadian law enforcement agency's cybercrime unit in a collaborative effort. To detect and extract forensically relevant information from large suspicious chat logs, a Word Net-based criminal information mining system. The system

processes a suspect's chat log to recognize a collection of cliques and topics in each clique's conversation.

According to [5] a new paradigm is suggested for cybercrime offences by feature extractions. Any unstructured cyber-crime report can be uploaded to produce the structure data via the TFID technique in this proposed system. This framework will subsequently report on the categorization and resolution of cyber-crime offences (particularly ID theft, hacking and copyright attacks) by their severity and incidence. Data preprocessing is a data mining technique that involves the transformation of unprocessed data into an understandable format. Raw data is often inadequate, incompatible and includes numerous errors and noisy data.

According to [6] a machine learning classifier created to find PHP code vulnerabilities for SQL injection. Using input validation and sanitization characteristics collected from source code files, classifier models were trained and evaluated using both conventional and deep learning-based machine learning techniques. A model developed using a convolutional neural network was validated 10 times (CNN). One of the most serious types of vulnerabilities that online applications are vulnerable to is SQL Injection (SQLI), which is the introduction of malicious code into SQL statements via web page input. In recent years, as there have been more and more online applications, SQLI has constantly been listed as one of the top 10 security concerns by the Open Web Application Security Project (OWASP).

According to [7] a strategy based on learning that creates abstractions of SQL Injection susceptible programmers from training datasets and groups them using hierarchical clustering. A fix proposal is produced when the test samples are matched with a group of related samples. The language used to communicate with relational databases is known as structured query language (SQL). There are several SQL statements used to carry out the interaction. The SQL injection attack (SQLI) takes advantage of SQL statement inputs. The attacks are often carried out by introducing special characters or keywords into SQL queries.

According to [8] on the cryptographic algorithms used in edge computing and provide a fresh approach to investigate the additional details discovered by the conventional LSM-based collision attack on masked AES. By using this information, a collision may be found quickly rather than by thoroughly scanning the plaintexts. To explain our solution and do tests to confirm its effectiveness, we used AES implemented with masks, which is often used in edge computing devices.

According to [9] a novel sort of collision attack using leakages from linear layers that is capable of destroying masking schemes with uniformly distributed random masks. The attack focuses on three prominent AES implementations in edge computing. Additionally, a brand-new, very effective collision strategy with broad application is suggested and implemented to masked linear layers and masked S-boxes. With sufficient off-line search, it may perform at a level comparable to second-order power analysis, which considerably enhances known collision attacks.

According to [10] as a result, phishers may quickly initiate phishing assaults using QR codes. The phishing attempts that use the QR code as a vector right now are analyzed and categorized in

this article. The most current defenses against these threats are also reviewed. Additionally, it is discovered that the present defenses are inadequate and have trouble dealing with problems like barcode-in-barcode assaults, high overhead solutions, and limited data space in the code. The effort done to identify phishing in QR codes has not kept pace with that of email and online phishing. This research aims to shed light on the latest QR code-based phishing attempts and the suggested defenses against them. Recent phishing techniques that use QR codes have led to the observation that real QR codes are still being manipulated. It is simple to conduct such an attack by covertly changing or manipulating the current QR codes to change the link's final destination. According to reports, it is possible to manipulate QR codes to produce barcodes inside barcodes. Given that QR codes can tolerate a certain amount of data mistake, a portion of the code may be altered to include a different barcode. Such manipulation may be used to launch an attack against certain scanners that are vulnerable enough to read the encoded barcode.

Attack Detection using Forensic Investigation in Cloud using VM

Collision attack

A collision attack simply refers to the process by which an attacker utilizes one of these clashes to undermine the security that the hash was supposed to provide.

Collision attack is an attack where the malicious user (i.e untrusted user) gives the wrong or misleading password about the cloud service provided by the cloud provider. In this malicious user means the user or an attacker who is never registered to the cloud service or the one who is not used the cloud service provided by the cloud provider. In this project we prevent the collision attack by avoiding the misleading password from unregistered users. Thus we check whether the user who is giving the password is registered or not. If the person giving the password is registered then he/she is called as the cloud consumer and their comments are accepted and displayed to the new user.

Password login: A collision is a condition whereby two messages, let say $D1$ and $D2$, after applying the hash value, then $H(D1) = H(D2)$. A collision can always be found using Brute Force algorithm, however it is computationally difficult. There are two types of collisions, the strong collision and weak collision.

Multiple malicious requestors collude to compromise the victim's privacy by coordinately launching queries on nodes' friendship and sharing their query results. These malicious requestor accounts can be created and manipulated by either a single or multiple real human attacker(s).

SQL injection

SQL injection is another form of attack that occurs when SQL queries are made with user input text inserted into the query string. QR code readers are subject to data injection into their

structured objects when they attempt to interpret the data of a QR code. A malicious party can create a QR code that injects arbitrary strings into a user's data structures potentially causing harm to the user.

The following things might result from SQL Injection

- Hacking other person's account.
- Changing the system's sensitive data.
- Deleting system's sensitive data.
- The user can log in to the application as another user, even as an administrator.
- Users can view private information belonging to other users e.g., details of the other users' profiles, transaction details, etc.
- The user could change application configuration information and the data of the other users.
- The user could modify the structure of the database; even delete tables in the application database.
- The user can take control of the database server and execute commands on it at will.

SQL Injection Attack

There are different types of SQLi attacks such as error-based, Boolean-based, time-based, and out-of-band SQLi. When exploiting an error-based SQLi vulnerability, attackers can retrieve information such as table names and content from visible database errors using the following queries:

'and(select+1+from(select+count(*),floor(rand(0)*2)from+
information schema.tables+group+by+2)a)–+
id=1+and(select 1 FROM(select count(*),concat((select (select
concat(database())) FROM information schema.tables LIMIT 0,1),
floor(rand(0)*2))x FROM information schema.tables GROUP BY x)a)

Attackers can test for web application vulnerabilities to SQLi by inserting a condition into an SQL query. If the page loads as usual it indicates that the page is prone to attack. The following query is an example:

id=1+AND+1=1

This type of query is considered as Boolean-based SQL Injection.

In some cases, even though a vulnerable SQL query does not have any visible effect on the output of the page, it may still be possible to extract information from an underlying database. Hackers determine this by instructing the database to wait (sleep) for a stated amount of time before responding. If the page is not vulnerable, it will load quickly; if it is vulnerable it will take longer than usual to load. The SQL syntax can be similar to the one used in the Boolean-based SQLi vulnerability. However, to set a measurable sleep time, the 'true' function is changed to something that takes some time to execute, such as instructing the database to sleep for three seconds, as follows:

id=1+AND+IF(version()+LIKE+'5%',sleep(3),false)

Sometimes the only way an attacker can retrieve information from a database is to use out-of-band techniques. Usually, these attacks involve sending the data directly from the database server to a machine controlled by the attacker. Attackers may use this method if an injection does not occur directly after the supplied data is inserted, but at a later point in time:
id=1+AND+(SELECT+LOAD FILE(concat('\\\\',(SELECT @@version),'abc.com\\')))
' UNION SELECT username '~' ' password FROM users—
This uses the double-pipe sequence —— which is a string concatenation operator in Oracle. The injected query concatenates together the values of the username and password fields, separated by the '~' character. The results from the query will let you read all of the usernames and passwords. Finally, the following query can drop important tables from the database such as table containing credentials or credit card information.

Digital Forensic Investigation Life Cycle

From the digital forensic definition, digital forensic investigation process involves many steps as follow

Identification: It is involved in two key phases: identification of crime and identification of digital evidence.

Collection: In this phase, an examiner gathers digital evidence from the crime scene for using in the next examination phase.

Extraction: In the extraction phase, the digital investigator extracts digital evidence from various types of devices such as cell phone, hard disk, and e-mail.

Analysis: In this phase, the examiner interprets and correlates the extracted digital evidence to come to a summary, which can prove or disprove criminal accusations.

Examination: In the examination phase, the investigator extracts and inspects the data and their characteristics.

Report: In this process, the investigator and examiner make a prepared report to represent his/her findings from forensic analysis of crime evidence. This report should be suitable enough to present in the court of law.

Limitations of this study

- This paper focuses on SQL injections attack detection and prevention utilized on a synthetic dataset.
- Low accuracy rate for MiM and zero-day attacks
- It works based on historical data; it may hard to detect an actual malicious user in the runtime environment
- Improved accuracy and protection against insider attacks.

Conclusion

In this research, we offer a new method of employing virtual machine proof to efficiently enable cloud-based digital forensics. This study presents a research system with the overarching goal of providing a model that effectively covers the many facets of integrated cybercrime occurrences using a single, unified framework. The first step is to efficiently find cybercrime traits from the dataset, and the second is to classify specific crimes based on those features. Cybercrime statistics are also analyzed, and predictions made as to which types of cybercrime will be most common in a certain year and geographical region. This technique improves cloud performance in terms of space and time by keeping malicious VM snapshots, and it does so by integrating an intrusion detection mechanism into the VM to identify malicious VMs. The proposed technique improves cloud performance by saving snapshots of potentially malicious virtual machines.

References

[1] Yeboah-Ofori, Abel, Ezer Yeboah-Boateng, and Herbert Gustav Yankson. "Relativism Digital Forensics Investigations Model: A Case for the Emerging Economies." 2019 International Conference on Cyber Security and Internet of Things (ICSIoT). IEEE, 2019.

[2] Aldaej, Abdulaziz. "Enhancing cyber security in modern internet of things (iot) using intrusion prevention algorithm for iot (ipai)." IEEE Access (2019).

[3] Singh, Kumar Shanu, Annie Irfan, and Neelam Dayal. "Cyber Forensics and Comparative Analysis of Digital Forensic Investigation Frameworks." 2019 4th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2019.

[4] Iqbal, Farkhund, et al. "Wordnet-based criminal networks mining for cybercrime investigation." IEEE Access 7 (2019): 22740-22755.

[5] Sudha, T. Satya, and Ch Rupa. "Analysis and Evaluation of Integrated Cyber Crime Offences." 2019 Innovations in Power and Advanced Computing Technologies (i-PACT). Vol. 1. IEEE, 2019.

[6] Zhang, Kevin. "A machine learning based approach to identify SQL injection vulnerabilities." 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, 2019.

[7] Siddiq, Mohammed Latif, et al. "SQLIFIX: Learning Based Approach to Fix SQL Injection Vulnerabilities in Source Code." 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). IEEE, 2021.

[8] Ding, Yaoling, et al. "Adaptive chosen-plaintext collision attack on masked AES in edge computing." IEEE Access 7 (2019): 63217-63229.

[9] Niu, Yongchuan, et al. "An efficient collision power attack on AES encryption in edge computing." IEEE Access 7 (2019): 18734-18748.

[10] Yong, Kelvin SC, Kang Leng Chiew, and Choon Lin Tan. "A survey of the QR code phishing: the current attacks and countermeasures." 2019 7th International Conference on Smart Computing & Communications (ICSCC). IEEE, 2019.