NGSIEM Based APT Attack Analysis System

Yanfei Li\textsuperscript{a}, Rui Wang\textsuperscript{b}, Haiwei Li\textsuperscript{c}

Information security department FRI Beijing, China

\textsuperscript{a}liyanfei_byr@163.com, \textsuperscript{b}ruiwang114@163.com, \textsuperscript{c}304055019@qq.com

Abstract. APT attack has become the most serious security threat, attackers may use social engineering methods and a large number of 0-day vulnerability attacks to steal or tamper the core data of the target. If the target does not have the ability of real-time attack detection and defense, once the system is intruded, it will suffer serious economic and business losses. The author sums up the typical characteristics and life cycle of APT attacks, which explains the common attack channels and critical steps of APT attacks, and then describes the technical difficulties and challenges in analyzing APT attacks. To solve the problems, an APT attack analysis system based on NGSIEM is proposed, which can implement normalization and complex algorithm processing on logs and alarms collected from server devices and security devices of multiple layers of the business system, as well as threat intelligence acquired from the intelligence agency.

1. Introduction

With the increasing severity of cyber-attacks, more and more organizations have been suffered from APT (Advanced Persistent Threat) attacks. Although large numbers of security devices have already been deployed to enhance security capabilities, these organizations still suffer persistent threats from APT attacks, which may lasted for several months or even longer.

Aiming to find effective analysis techniques to enhance the security of important business systems, and promptly discover and predict APT attacks, an APT attack analysis system based on next generation security information and event management (abbr. NGSIEM) was proposed. NGSIEM can be used to manage massive security data collected from all security devices and audit applications, and can also provide a significant advance on reducing time for correlating, consolidating and contextualizing diverse security event information in each key step of APT attacks, and correlating long-term historical data for forensic purposes [1].

2. Literature Survey

2.1. Characters of APT attack

When APT attackers find the valuable targets, they would exploit application vulnerabilities and launch persistent attack until they can penetrate into key region of the systems, then maintaining some control channels, and finally steal or modify the core data will be much easier. It is different from all the security vulnerabilities, and main characters are as follows:
Latency: Attackers usually hide in a target for months or even more than a year, and collect a large amount of accurate information about the target system, to comprehensively grasp the target’s information.

Targeted: After collect accurate information of the target, attackers would try to find software vulnerabilities and intrude the system to gain access authority by constructing special code, or sending malicious links, mails and other programs to the target.

Persistence: If the attack processes were not detected, attackers would constantly try various attack approaches. If blocked, other ways would be adopted to launch attacks, so some attacks would last even for several years.

2.2. Life cycle of APT attack
The life cycle of APT attack can be divided into five stages [2]:

First stage: directional intelligence collection. Attackers purposefully gather information to understand the overall network structure and locate important roles which can access data resources or can be served as springboards.

Second stage: single point breakthrough. After gathering enough intelligence information, attackers would use social engineering methods and 0-day vulnerability attacks. By social engineering attacks, attackers can send mail to employees of the target company, trick them to open malicious attachments and sending malicious code. By 0-day vulnerability attacks, attackers can upload a trojan horse on a website frequently visited by employees of the target company. When an employee was accessing the site, the malicious code would be downloaded and installed on the employee's device. The above two types of attacks are the most common methods to start the effective infection and access core resources.

Third stage: control channel construction. After controlled the terminal device, attackers would create command control channel from the controlled terminal to the control server to obtain further attack instructions. To maintain the normal access of the channel, attackers would continuously enhance the access rights to obtain more system operation privileges that make the attack behavior hard detected.

Fourth stage: internal horizontal penetration. In general, attackers would take precedence to capture employee terminals as a springboard for attack, using password eavesdropping and vulnerability attacks to penetrate more terminals and servers within the system to gain control of other servers which containing important assets within the organization.

Last stage: data collection and theft. During the attack, attackers would compress, encrypt and package the important data resources collected from the servers, and then send back the data through the control channel. If successfully, attackers would delete all intrusion traces, attack traces, data records, etc.

2.3. Capabilities of NGSIEM
SIEM is the security solution which gave ability to monitor all security events happening inside the system and represent logs in a well-mannered form. NGSIEM has upgraded SIEM functions and its new capabilities are as follows:

Big Data: With big data technologies deployed to provide storage, analysis and process capability for massive events, NGSIEM can further enhance depth and breadth of event collection and also extend storage time limit of original event metadata.

Muti-Structured Data Support: Through using index, content analysis, semantic analysis and other technologies, unstructured and semi-structured security event related data can be analyzed and processed, and the traditional paradigmatic event access mode is extended.

Advanced BI: NGSIEM can expand its capabilities of non-event-based and feature-based analysis, and discover abnormal behaviors with BI presentation technology, by analyzing trends, categorizing behaviors, and conducting other complex models.
Mutli Security Event Support: To expand the coverage of security events from pure infrastructure events to data security events, application security events and even business security events, NGSIEM is gradually upward security to an anti-fraud, IT management, and even business analysis platform.

More Rapid Real-Time Analysis and Processing Capability: Hadoop and other technologies actually have a series of problems, such as single thread, not real-time, etc. Through integrating spark engine, NGSIEM can provide faster real-time analysis and processing capability.

3. Technical Challenges and Issues
APT attacks have changed the cyber threat from random attack to group attack, which is purposeful, organized and premeditated. New challenges have brought to the traditional detection technology. The main issues are as bellows:

Firstly, APT attackers often use 0-day vulnerabilities to obtain privileges, and carry out remote control through unknown and special trojan horses, which cannot captured by traditional detection devices based on feature matching.

Secondly, APT attackers are skilled in building secret and encrypt channel, or forging legal signature to avoid malicious code files being recognized, which brings great difficulties to traditional detection based on signature.

Thirdly, APT attacks usually last and hide for a long time and can be divided into several steps. The traditional detection method is based on real-time detection of a single point in time, and it is difficult to track such a long span of attacks effectively.

Finally, any cyber exposure point may be the target of APT attacks, and the means of attack are various in different situations, therefore, the attack processes are very difficult to analyze and reconstruct.

Although there are many challenges mentioned above, about 86% of publicly exposed APT attacks are monitored and recorded [3], so APT attackers will leave traces on different security devices during their attack processes, and the detection accuracy of APT attacks based on massive data analysis is much higher.

In order to cover a whole APT attack process, full traffic auditing is needed to store and analyze to identify whether there is potential attack. Due to such large scale of data, big data technologies and platforms are needed to analyze all kinds of possible APT events, and then organize the events in series according to certain correlation.

4. Detailed Design of System
The system is composed of three functional modules.

Security logs are gathered by log collectors, such as NGSIEM.

Security alarms are collected by terminal-oriented and traffic-oriented anomaly detection, as well as threat intelligence acquisition.

Association analysis and machine learning algorithm analysis of massive data are conducted to further discover APT attacks.

4.1. Data Collection and Storage
In the module, in order to collect security information by full traffic auditing, three typical audit logs must be collected and stored, including system logs, application logs, and other network device logs.

NGSIEM is used as a log collector to monitor and collect massive homogeneous and heterogeneous logs from various devices. When these logs have been collected, NGSIEM is also used to convert raw data into events and store events in an indexing engine for retrieval communication.

4.2. Threat Detection
In the module, three types of anomaly alarms should be detected to spot the multiple layer threats of the entire system [4], which are as follows:
4.2.1. Terminal-oriented anomaly detection. To monitor the running state of the application system at all times, protection software and probes must be installed on the system-deploying terminals. Through RASP technology, the traffic, context and behavior of the application system are continuously monitored to identify and defend against known and unknown threats. In addition, by using probes in the kernel layer, the behavior of the system can be continually learning, and the abnormal behavior in the kernel system can be effectively detected.

4.2.2. Traffic-oriented anomaly detection. Traffic-oriented anomaly detection can capture a large number of anomaly behaviors in the whole network, and also discover a large number of attack traces and malicious applications. Therefore, the entire detection system should use IDS, IPS, firewalls, honeypots, and other security hardware devices to locate and identify malicious traffic and abnormal access behavior by real-time network traffic monitoring.

4.2.3. Threat intelligence acquisition. The above detection is analyzed from the perspective of goals and means, another important information is about the attackers and their social network security threats. Therefore, the system should have the ability of threat intelligence acquisition, and threat intelligence is mainly divided into the following two categories.

The abnormal behavior violating the normal behavior pattern which is mining from the information content and relationship of social network [5].

The social attributes of suspicious attackers from social network data which can provide guidance for the trace ability of attack events and the identification of attack intentions [6].

4.3. Big data Analysis

In the module, analyzing above data in different business scenarios is the central section, but the difficulty also lies in this. Data analysis mainly requires security experts who are familiar with the business to give a large number of business scenarios and business rules with practical significance. It is not difficult to abstract the rich security experience into business rules and accumulate them continuously, but an interactive analysis tool which can help experts manage and test the rules on such massive security data is needed.

NGSIEM, which is a business analysis platform with machine learning toolkit, can be used to discover various kinds of APT attack events by defining the simple analysis models. The APT event features are shown in Table I, which are classified according to the life cycle of an APT attack.

It is necessary to conduct machine learning algorithm analysis to discover hidden attack patterns in above events, and form filtering or association rules. Non-relational data management technology represented by Map Reduce [7] technology has obvious advantages on big data association analysis, including time series analysis, attack process analysis and so on.

Time series analysis [8] uses the method of mathematical statistics to process the data series of each event in time sequence, in order to predict the future attack trend. It is meaningful to study the change of local features of time series and define the local patterns that identify the occurrence of major events. Moreover, the correct recognition of patterns is helpful for security experts to analyze attack scenarios and use similarity function between events to construct attack sequence set.

Attack process analysis [9] makes association analysis of attack events in order. This kind of method uses expert experience to construct attack process model by making causal association rules based on the causal logic among APT attack steps. Attack intent can be identified by matching the detected attack with attack process model, which will provide basis for attack detection in the next stage. Therefore, it is convenient to match attack context to determine the attack type of each attack scenery, so that the security experts can effectively predict all possible attack processes.
Table. 1  APT Attack Event Features

| Attack Stage | Attack event | APT Attack Event Features                        | Rule description                                                                 |
|--------------|--------------|--------------------------------------------------|----------------------------------------------------------------------------------|
| Stage Two    | Port scan    | In a certain time range, the number of ports receiving requests exceeds the threshold. |
| Stage Two    | Cipher blasting | In a certain time range, the number of failed requests for a single IP exceeds the threshold. |
| Stage Two    | Malicious mail | The target host receives mail which contains attachments, and the email address is similar to but not its common contacts. |
| Stage Three  | Malicious script upload | Malicious script is uploaded to the catalog of target website. |
| Stage Three  | Unauthorized process execution | The installation program is installed on the target host, and executed in an unauthorized state. |
| Stage Four   | Remote control | The target host opens remote access control.      |                                                                                  |
| Stage Four   | File system access | The file catalog of the target host is traversed orderly within a certain time. |
| Stage Four   | Database system access | The database access service of the target host is opened, and the data items are regularly traversed within a certain time. |
| Stage Five   | Web page tamper | The web page is tampered.                        |                                                                                  |
| Stage Five   | Suspected data theft | In a certain time range, the network traffic exceeds the threshold. |

5. Conclusion

This paper provides an explanation of the research on the characteristics and life cycle of APT attack, and explains the way to deal with these issues and challenges. An efficient APT analysis system developed based on NGSIEM, which can incorporate all kinds of security data collected in various ways, and provide a more friendly way of data management and analysis for security experts. The deep association analysis of detection data using large data technology is also proposed to comprehensively analyze whether the target system has the risk of APT attack.

In the future, in order to improve efficiency, more attention should be paid to the study of frequent itemsets mining algorithms [10]. The algorithms, like big data mining based on similarity discovery, frequent itemsets mining based on constraints, and incremental mining of frequent itemsets, will be integrated with the system to provide analysis convenience for security experts.

References

[1] Big Data Analytics for Security Intelligence, Cloud Security Alliance, September 2013.
[2] ZHANG Yu, PAN Xiaoming, LIU Qingzhong, CAO Junkuo, LUO Ziqiang. APT attacks and defenses. Journal of Tsinghua University (Science and Technology), 2017, 57(11): 1127-1133.
[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] VRIES, J D, HOOGSTRAATEN H, BERG JV D, et al. Systems for detecting advanced persistent threats cybersecurity: a development roadmap using intelligent data analysis[C]. Proceedings of the 2012 International Conference on Cyber Security. Washington, USA: IEEE Computer Society, 2012: 54-61.
[5] CUI X, LIU C G, CHENG X Q. APT analysis and consideration of big data calculation. China Information Security, 2014, No.01, pp. 102-104.
[5] LINDAMOOD J, HEATHERLY R, KANTARCI OGLU M, et al. Inferring private information using social network data[A]. International Conference on World Wide Web[C]. Madrid, Spain, 2009.123-130.

[6] LI Y, WANG X Y, WANG K, et al. Security relationship based on social network[J]. Journal of Computer Research and Development, 2012, 49(suppl): 124-130.

[7] CHEN X Q, JIN X L, WANG Y Z, et al. Survey of big data system and analysis technology[J]. Journal of Software, 2014, 25(9): 1889-1908.

[8] QIN X, LEE W. Statistical causality of INFOSEC alert data[C]/ Re-cent Advances in Intrusion Detection 2003. Berlin, 2003: 73-93.

[9] CHEN X J, FANG B X, TAN Q F, et al. Inferring attack intent of malicious insider based on probabilistic attack graph model [J]. Chinese Journal of Computers, 2014, 37(1): 62-72.

[10] WANG L, FENG L, WU M. An efficient algorithm of frequent itemset mining on uncertain dataset[J]. Journal of Computers, 2013, 8(6): 1417-1426