A New Detection Method of Network APT Based on Big Data Analysis

Min Li

College of Electronic Information Engineering, Chongqing Technology and Business Institute, Chongqing, 400052, China

email:503783334@qq.com

Keywords: Network APT, Detection method, Big data analysis

Abstract: APT (Advanced Persistent Threat) is a persistent and complex network attack pointing to clear targets. The concept became one of the hot topics in the information security industry after Google admitted to severe hacker attacks in 2010. Big data analysis provides a good data base for APT detection. This paper analyzes the general process of APT, and gives a new APT detection method based on big data analysis to provide some references for the relative researchers.

1. Introduction

Network security is facing unprecedented challenges. This is mainly due to new attacks and threats of organized, specific goals and long duration. We call it APT (Advanced Persistent Threat). In the APT attack defense work, attack detection is the premise and basis of security protection and reinforcement, and it is also the most difficult part of APT attack defense. Therefore, detection technology has become the research hotspot in the field of APT attack defense. However, from the typical case, APT attack has strong concealment ability and pertinence, and the traditional detection equipment is mostly helpless in the face of APT attack. As the security data in the network has the characteristics of large volume, diverse sources, rapid expansion speed and low value density, the security detection technology based on network big data analysis is gradually emerging in recent years. The technology can realize the deep association analysis of massive network security data, and intelligently correlate multiple types of security events in a wide time window, so it has obvious advantages in detecting APT attacks. APT attack is a new attack and threat which is organized, targeted, concealed, destructive and long lasting. Its main characteristics are various means, clear goals and long duration. At present, APT attack has become a hot topic in the field of Internet security, and continues to heat up. APT attack is a new attack and threat which is organized, targeted, concealed, destructive and long lasting. Its main characteristics are various means, clear goals and long duration. At present, APT attack has become a hot topic in the field of Internet security, and continues to heat up. In the era of big data, due to the huge volume of data, widely distributed, it brings new challenges to security issues. There is a corresponding and parallel relationship between the real space and the data space. Any activity, interaction and behavior in the real space have corresponding performance in the data space. Therefore, the means and solutions existing in the data space can affect the real space. The role of data space is ubiquitous, which is also the value of big data.

2. General Process of APT

2.1 Search Stage

Compared with the ordinary network attack, the APT attack has obvious difference in the depth and breadth of information search. APT attackers spend a lot of time and effort searching for information about the target system. They will understand the background of the enterprise, corporate culture, personnel organizations, but also collect the target system network structure, business

systems, application versions and other information. Then the attacker will plan well, to help achieve the target recognition system, personnel information collection, development or purchase of attack tools, APT attacks may use special exploit tools, password guessing and other tools, penetration testing tools. The main task of this stage is to collect information, the attacker plan, at this stage we can rely on the security threat detection and warning system, identify the attacker to the enterprise network sniffer, scanning behavior, so advance prevention; to strengthen the safety management of information systems, such as regular safety inspection and reinforcement, as little as possible to increase the initial exposure information system. The difficulty of the attack. We can regularly train staff awareness of safety, improve staff awareness of security.

2.2 Ingress Stage

The attacker will attempt to break through until the breach is found and the first computer controlling the intranet is controlled. Malicious files are PDF and Word documents with malicious code sent to employees by mail and IM software. Malicious links are sending URL links with malicious code to employees by mail, IM software and other forms to entice employees to click. Website vulnerability is the use of loopholes in the website system, such as SQL injection, file upload, remote overflow, etc., to control the website server as a springboard, infiltration and attack within the enterprise. This is one of the most convenient attacks, from the underground black market directly to buy computers that have been compromised by other hackers inside the enterprise. The entry point APT attack exploits vulnerabilities, uses a large amount of information gathered to attack the established target host, sends malicious programs to the target host, and then induces the user to run malicious programs. Another method is to send a malicious object to a given target. When the mail is opened, it will download the malicious program and execute the program itself. The computer will try to access the internal attackers at this stage as the first goal of implementation of intrusion behavior, at this stage we can rely on the aggressive behavior of security threat detection and warning system to identify ongoing security policy; rational allocation of intrusion prevention system, such as firewall products, blocking the conventional attack attempts to raise vigilance, avoid the attacker behavior; the social engineering trick; once found attacks start event disposal and emergency response procedures.

2.3 Infiltration Stage

Attackers use the computer that has been controlled as a springboard, through the remote control, to penetrate the enterprise intranet, looking for valuable data, like the entry stage, this stage will test the attacker's patience, technology, means. The penetration stage is like the entry stage, and attackers try different attack techniques and attack methods to invade the target system. In addition, the threat monitoring and security awareness are also the key points to be strengthened, such as the rational planning of security domain, the security audit of system accounts, the system account and authority management, and the optimization of system security policies. APT attacks have sensitive information through search is important to determine whether the computer, the computer will find the target through the network communication protocol and establish communication, establish a network firewall through secret channel between the controlled computer and the server, and confirm the successful invasion of computer and server communication.

2.4 Harvest Stage

The attacker made preliminary success, they will build a hidden data transmission channel, will have access to confidential data transfer out, the stage name harvest stage, but there is no time limit, as the initiator of APT attack and common attack than extreme greed, if not to be found aggressive behavior, often do not stop, continue trying to steal sensitive data and new confidential information. At this stage, the attacker will try to transfer the acquired confidential data to the external network of the enterprise, so the detection of sensitive traffic and illegal connections becomes particularly important. We use channels to find important information, establish the target system, APT in the target network through the search of ancient store sensitive information is important to determine the

old computer, after finding the target computer using the algorithm contains specific skills and tools, make use of the attacker's permissions to a higher level, so that the attacker can easily access and control the target computer. APT attacks cannot be effectively detected and protected by a single security product and security technology. Enterprises can only defend against the threat of APT attacks by establishing a deep protection system combined with security technology and security management. In addition, the general process of APT attack and defense, threat detection throughout, because only timely detection of APT attacks, we can in the first time to prevent the network attacks continue to deteriorate, and then targeted to improve the enterprise security protection system.

3. New Detection Method of Network APT Based on Big Data Analysis

3.1 Refine Data Comprehensively

We refine the huge data, filter the data in the full flow, delete the data without any correlation with the attack behavior, and keep the relevant data traffic at the same time, which can release more space. Select, delete no correlation to achieve data traffic, you can rely on the third-party detection alarm device, you can also use the full data traffic anomaly detection technology. In the process of social network security behavior mining, although the data source is non-sensitive data, it is still possible to tap the privacy of normal users, causing legal disputes for security behavior diggers, and thus affecting the normal work of network security detection. Privacy preserving data mining is one of the effective ways to solve this problem, however, privacy preserving data mining method based on interference can reduce the availability of mining algorithm, influence accuracy of attack detection; privacy protection encryption of data mining method based on high cost, not easy to deploy in social networks. Therefore, it is a big challenge to protect the privacy information of normal users in the process of social network security behavior mining. Establishing the feedback mechanism of network security detection, and making full use of the information of suspicious attackers to guide the mining work is the development direction of social network security behavior mining. In addition, although the security incidents in a social network can provide guidance for network attack detection, but the web crawler to run user nodes using an initial drive, considering the APT attack process of typical mostly use social engineering and social network analysis methods were used to collect sensitive information, so the operation process of the web crawler may be APT the attackers found. Designing high reliability and concealment crawler is the development direction of social network security data collection. We develop a practical and strong operational scheme defending against APT attacks imminent from a technical perspective, APT has the latent features. Many attacks are likely to wait for a long time in the target network environment. Through the effective collection of information continues to attack the target, the persecution of as a result, to strong attack, will make the target unable to fight back. The analysis showed that the" big data "is to have certain rules, through the transverse or longitudinal data association, matching process can be created with their own defense system in the network environment, investigation and the comparison of network anomaly data, it can be judged that information in the network environment with the exception information is the nature of the attack, if found the potential attack line You can quickly target the attacker.

3.2 Give Alarm Accurately

We give accurate alerts to detected attacks. We will continue to analyze the data related to the aggressive behavior, and make further accurate alarm. Malicious code under the condition of big data anomaly detection requirements of feature extraction process automatically, quickly and effectively, and the dynamic feature extraction method with sample low coverage, feature extraction speed is slow, so the static analysis method of expression characteristics more suitable for big data conditions, however, because the APT attack is a high level of professional attackers take regardless of cost of attack, so the vast majority of APT malicious code are used in encryption, code obfuscation and other covert means, resulting in the static analysis method can effectively extract feature code. Therefore, how to solve the timeliness problem of dynamic feature extraction and the static feature extraction

cannot identify the hidden code problem is a big challenge. These collected samples are in a certain period, because the network status and APT attacks of information are dynamic changes, so the use of a learning results repeated detection, the effect will be very poor. At present, the main method of malicious code feature extraction for APT attack is sandbox analysis technology. How to combine the static features of code to reduce the time and space cost of sandbox analysis is the development direction of malicious code feature extraction research field. Due to the strong persistence and penetration of APT attacks, the network traffic anomalies caused by APT attacks usually have obvious delay, which leads to the invalidation of the anomaly detection methods in narrow time windows. The traditional traffic anomaly detection technology in the massive network flow data as input, so the wide time window within the network flow data volume is too large, not only makes the data analysis cost increased sharply, and the data were too large easy to produce many false anomaly detection and lack of guidance, which led to the detection result is meaningless.

3.3 Construct Scene Rapidly

We will make a precise analysis of the relevance of the last step of the alarm, make sure what semantic relationship exists between them, extract the isolated attack alarm, build a whole blood, integrated attack scenarios. Anomaly identification scheme based on deep level protocol parsing. We can carefully check what kind of agreement is found, where a data is abnormal, and stop detecting until it finds out its abnormal point. An attack tracing scheme; a network object that has been extracted can reconstruct all the suspicious content in a single time. By rearranging these events, we can quickly find the source of the attack. Considering the inherent characteristics of massive data analysis, blind large-scale data analysis of massive wide area network data is often not only unable to effectively detect attacks, but will result in many irrelevant detections results due to the lack of guidance in the analysis process. In the wide time domain, the number of security events extracted by data association analysis is large, and the types of security events are diverse, and there are a lot of false alarm information. Therefore, it is necessary to extract the real attack information from many security events through the security association method. Based on this, this layer uses network intrusion security events, malicious code and user behavior security events inherent association of several attributes, the use of security event correlation analysis method to identify the attack occurred time, place, attack type and strength and other information. Therefore, the initial detection results in narrow space domain can guide the next step of detection process, and the wide range detection of key time segments of key nodes is targeted. This feedback detection method can improve the detection rate while improving the timeliness of detection, so it can be scientific and practical detection of APT.

4. Conclusion

APT attack method is complex, diverse forms of expression. This will pose a new threat to network information security, but at the same time, big data also provides a more effective way for us to chase APT attacks and do a good job of security defense. The complexity of the network environment, relying on unilateral strength to detect APT attacks is not reliable. To make up for the shortcomings of the traditional real-time monitoring, we need to monitor and analyze the large data of long time and full flow.

Acknowledgement

The paper is the result of Project Supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. kj1603810).

References

[1] Fu Yu, Li Hongcheng, Wu Xiaoping, et al. Detecting APT attacks: a survey from the perspective of big data analysis [J]. Journal on Communications, 2015, 36(11): 1-14.

[2] Zhang Hao, Wang Lina, Tan Cheng. Review of Defense Methods Against Advanced Persistent Threat in Cloud Environment [J]. Computer Science, 2016, 43(3): 1-7+43.

[3] Du Yuejin, Zhai Lidong, LiYue, et al. Security Architecture to Deal with APT Attacks Abnormal Discovery [J]. Journal of Computer Research and Development, 2014, 51(7): 1633-1645.

[4] Tan Ren, Yin Xiaochuan, Lian Zhe, et al. Hierarchical representation model of APT attack [J]. Journal of Computer Applications, 2017, 37(9): 2551-2556.