Hindawi
Security and Communication Networks
Volume 2017, Article ID 7536381, 9 pages
https://doi.org/10.1155/2017/7536381

Research Article

Constructing APT Attack Scenarios Based on Intrusion Kill Chain and Fuzzy Clustering

Ru Zhang, Yanyu Huo, Jianyi Liu, and Fangyu Weng

Information Secure Center, Beijing University of Posts and Telecommunications, 10 West Tucheng Road, Haidian District, Beijing, China

Correspondence should be addressed to Ru Zhang; zhangru@bupt.edu.cn

Received 28 September 2017; Accepted 4 December 2017; Published 27 December 2017

Academic Editor: Zhenxing Qian

The APT attack on the Internet is becoming more serious, and most of intrusion detection systems can only generate alarms to some steps of APT attack and cannot identify the pattern of the APT attack. To detect APT attack, many researchers established attack models and then correlated IDS logs with the attack models. However, the accuracy of detection deeply relied on the integrity of models. In this paper, we propose a new method to construct APT attack scenarios by mining IDS security logs. These APT attack scenarios can be further used for the APT detection. First, we classify all the attack events by purpose of phase of the intrusion kill chain. Then we add the attack event dimension to fuzzy clustering, correlate IDS alarm logs with fuzzy clustering, and generate the attack sequence set. Next, we delete the bug attack sequences to clean the set. Finally, we use the nonaftereffect property of probability transfer matrix to construct attack scenarios by mining the attack sequence set. Experiments show that the proposed method can construct the APT attack scenarios by mining IDS alarm logs, and the constructed scenarios match the actual situation so that they can be used for APT attack detection.

1. Introduction

Nowadays, attacks on the network are becoming more and more complex, and, among them, APT attacks are increasingly frequent [1]. Unlike traditional attacks, APT attacks are not launched to interrupt services, but to steal intellectual property rights and sensitive data [2]. An APT attack has the stage and longevity characteristics and uncertain attack channel. Therefore, the Intrusion Detection System (hereinafter referred to as IDS) cannot detect an APT attack and can only generate alarms to certain steps in the attack. In 2012, Kabbah and Comodo companies' source codes were stolen [3]; in 2015, the OceanLotus Organization launched APT attacks on a number of essential institutions, including the Chinese government, certain research institutes, and maritime organizations in China [4]. Since then, APT attack has become a hot research topic. This paper focuses on how to correlate a large number of IDS security logs to dig out an APT attack scenario, and ultimately identifies an APT attack. Attack scenarios reflect the actual state of the network and can help defenders to take corresponding precautionary measures.

Correlating alarm logs is an important step to dig out attack scenarios. At present, researchers working on APT attack correlation built a full-scale attack model based on the phases of an APT attack and then correlated security logs with the attack model to generate the attack context. However, the establishment of APT attack model requires expert knowledge, and if the attack model is incomplete, some alarmed events will be unmatched and discarded, resulting in an incomplete attack route. In this paper, we propose a new method to solve this problem. We adopt fuzzy clustering correlation method to form clusters using multidimensional properties of alarm logs, so the correlated alarms are clustered to an APT attack route. Although each case is different, all APT attacks are phased, which conform to the feature of the intrusion kill chain model. According to this feature, we improve the fuzzy clustering algorithm by adding attack event property. We divide an APT attack process into several phases according to the intrusion kill chain model and categorize the attack events into different phases according to the characteristic of each phase, the behavior of each attack event, and the degree of harm. Then we compare the attack events of two alarms in the process of clustering, and if the attack event of latter alarm is in the subsequent attack phase relative to the attack event of previous one, then the correlation of the two alarms is stronger. The merits of taking the attack event as a cluster dimension are that it improved the correlation of alarms in an attack sequence, and there is no need to establish the attack model beforehand, and the alarm will not be lost because its event cannot match. Finally, we analyze the clustering results, combine the repeated attacks, delete the incomplete attack fragments, and then establish the probability transfer matrix to mining the attack scenario.

2. Literature Review

An APT attack is targeted, camouflaged, and phased, and it cannot be identified effectively with traditional detection technologies [5]. Friedberg et al. used the whitelist method to detect APT attacks. This method studied normal system behaviors and reported those operations different from system normal model, to find out Zero-day Threats [2]. Choi et al. used the extraction of normal behavior and anomaly patterns to detect the anomalies of APT attacks and proposed a method to detect anomalies by mining unknown anomaly patterns [6]. The APT attack model is often used in security log-based APT attack detection [7]; Tankard [8] established an APT attack model to monitor the network to discover the rules of actual attacking process. And Zhang et al. [3] constructed the attack tree model based on the intrusion kill chain and analyzed the attack logs to form the attack route to predict an APT attack. Three security researchers from Lock Martin first proposed the intrusion kill chain model on the ICIW Conference in March 2011 [9]. From the perspective of intrusion detection, this model decomposes the attacking process into 7 steps of reconnaissance, weaponization, delivery, exploitation, installation, command and control, and actions on objectives, and this model meets the phase characteristic of APT attack. APT attack detecting methods based on attacking model rely on expert knowledge predefining model. If the attack model is incomplete, attack scenario will be disrupted. If an attacker does not attack by well-defined rules and bypasses a phase in other ways, then a complete attack scenario cannot be constructed. Therefore, in this paper, fuzzy clustering is used in correlation to resolve these problems.

In the context of using fuzzy clustering to correlate alarms, the alarms are correlated to form an attack sequence by calculating the similarity between the alarms [10]. In terms of alarms correlation, most papers studied in general multistep attacks and made no adjustment according to different complex attacks. Feng et al. [11] used the correlation of the IP address in clustering, but the causation of two alarms is not just reflected in IP addresses. In this paper, we divide the attack events using the intrusion kill chain model and use multidimensional properties including the IP address, the attack event, and the time stamp in fuzzy clustering. This method resolves problems such as inability in constructing the complete attack scenario using expert knowledge, and loose coupling of clustering using single property. Finally, the attack sequences are fused by the transfer matrix, which avoids small-frequency attack sequences being omitted when using frequently occurring item sets.

3. Mining an APT Attack Scenario

IDS alarm log is a kind of log generated when attack operations occur. It shows security situation of the entire network.

Definition 1 (alarm logs). We represent IDS alarm log as alarms $= \{a_1, a_2, a_3, \ldots, a_n\}$, where a_i indicates the ith alarm and is a six-tuple:

$$a_i = (\text{timestamp}, \text{sIP}, \text{dIP}, \text{sPort}, \text{dPort}, \text{alarm\_event}). \quad (1)$$

The meaning of each attribute is shown in Table 1.

Definition 2 (attack sequence). An attack sequence is a sequence of IDS alarms that is produced by an attacking process. We represent the attack sequence as AS $= \{a_1, a_2, a_3, \ldots, a_n\}$, where all the alarms are listed in temporal order.

Definition 3 (attack scenario). The attack scenario shows the intrusion process of many different attack actions according to a certain time and logical sequence, which can be described in the form of graphs. Therefore, it can be said that the attack scenario consists of many single attack steps, which are the attack alarm information detected by the safety device.

3.1. The Entire Process. As is shown in Figure 1, there are four steps in the entire process:

(1) Data preprocessing: the IDS alarm log is normalized to the six-tuple format as in Definition 1 after a simple elimination of false positives in the data.

(2) Attack event classification: the APT attack is divided into several phases based on the intrusion kill chain model, and the attack events are classified according to the purpose of each phase and the behaviors of each attack event.

(3) Fuzzy clustering: the similarity function of each property used in fuzzy clustering is defined so that fuzzy clustering can be conducted. The attack sequence set ASS $= \{\text{AS}_1, \text{AS}_2, \ldots, \text{AS}_q\}$ is formed after fuzzy clustering, where each attack sequence $\text{AS}_i = \langle a_1, a_2, a_3, \ldots, a_n \rangle$ represents a possible APT attacking process, where a_i is an alarm.

(4) Attack scenario mining: we analyze all attack sequences generated after fuzzy clustering and delete isolated attack sequences without subsequent data transmission. A probability transfer matrix is established using multiple attack sequences where each row and each column represents an attack event. And finally, the probability transfer matrix is converted into a probabilistic attack scenario graph that can be used to identify APT attacks in the network.

Table 1: The meaning of each attribute.

The attribute of an alarm	Meaning
timestamp	The time when the attack occurred
sIP	The source IP address
dIP	The destination IP address
sPort	The source port
dPort	The destination port
alarm_event	The IDS alarm event

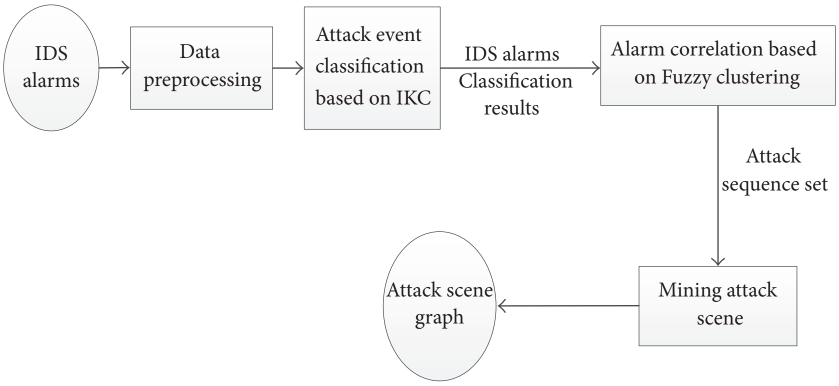

IDS alarms
Data preprocessing
Attack event classification based on IKC
IDS alarms Classification results
Alarm correlation based on Fuzzy clustering
Attack sequence set
Mining attack scene
Attack scene graph

Figure 1: The entire process.

3.2. Attack Events Classification Based on IKC. In some papers, the intrusion kill chain (IKC) model is widely used in constructing APT attack model. The attack event is an important property of an APT attack; therefore, the attack event dimension is included in fuzzy clustering, and all the attack events in the alarm log are classified by the purpose of each phase and the behavior of the attack event in the model, to calculate the correlation between two alarms in the fuzzy clustering. The two adjacent alarmed attack events are compared in clustering. If the latter alarmed attack event is in subsequent attack phase relative to the former alarmed attack event, the correction of the two alarms is higher.

We divide an APT attack into four phases based on the IKC model. Each phase has different purpose and different behavior.

(1) Information collection phase: it is the first step of an attack, including reconnaissance and information collecting, using some technical means such as scanning, probing, and social engineering.

(2) Intrusion phase: the attacker induces the target user to click on the phishing website or to download the malicious email attachment or install a backdoor through Trojan upload or loophole exploitation, to upgrade access permission to the target host.

(3) Latent expansion phase: the attacker maintains connection to the controlled host to obtain more valuable data and get ready for expansion. The attacker continues penetrating in the interior by using the host with permission as a stepping stone.

(4) Information theft phase: this is the confidential information transmission phase. The data will be transferred to the attacker's server after the attack has reached the host. The transport process often uses SSL or TLS secure transport protocol to encrypt data for camouflage. In addition to obtaining information, APT attackers can disrupt the facilities in the target network and interfere with the normal operation of the system.

We analyze the behavior and hazard of each attack event and classify all attack events into a certain phase. The classification process is shown in Figure 2.

3.3. Alarm Correlation Based on Fuzzy Clustering. Fuzzy clustering analysis generally constructs fuzzy matrix according to the property of the object and determines the clustering relationship according to the degree of membership. The properties of the alarm log are nonnumericand are typically measured in the following manner.

$x_i, x_j \in A$ where A is the alarm set, and the membership function of x_i and x_j in fuzzy clustering is defined as $S(x_i, x_j) = (\sum_{k=1}^{m} \alpha_k \cdot \delta(x_{ik}, x_{jk}))/m$, where m is the number of properties for an alarm, α_k is the weight of each property, and $\delta(x_{ik}, x_{jk})$ is the similarity function for each property, generated by the nature of property.

3.3.1. The Similarity Function. We define the similarity function of properties in fuzzy clustering according to different meanings of different properties. We define the similarity function of three properties including IP address, timestamp, and attack event as follows.

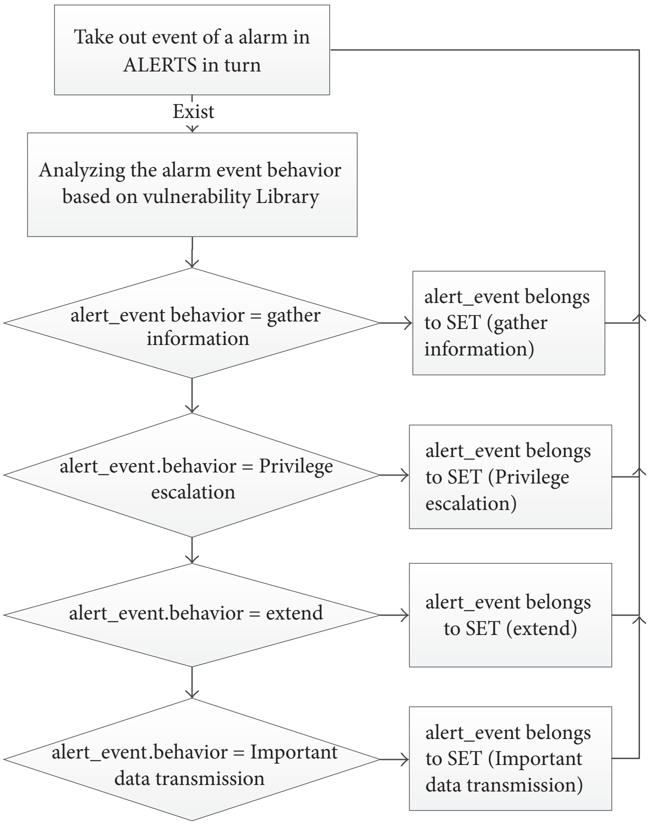

Take out event of a alarm in ALERTS in turn
Exist
Analyzing the alarm event behavior based on vulnerability Library
alert_event behavior = gather information
alert_event belongs to SET (gather information)
alert_event.behavior = Privilege escalation
alert_event belongs to SET (Privilege escalation)
alert_event.behavior = extend
alert_event belongs to SET (extend)
alert_event.behavior = Important data transmission
alert_event belongs to SET (Important data transmission)

FIGURE 2: The classification process.

a_i is an original alarm, and a_j is a classified alarm using fuzzy clustering. We use similarity of a_i and a_j to measure a_i's membership of the class containing a_j, that is, $F(a_i, a_j) = \delta_k F_k(a_i, a_j)$, where δ_k is the weight of each property, and k refers to alarm event, IP address, and timestamp three properties.

(1) The Attack Event Similarity Function. In terms of the attack event dimension, the similarity function of a_i and a_j to an attack sequence is as follows:

$$F_{\text{alarm\_event}}\left(a_i, a_j\right) = \begin{cases} 1, & \Delta\alpha = 0 \text{ or } 1 \\ e^{-(\Delta\alpha - 1)}, & \Delta\alpha > 1 \\ 0, & \text{else}, \end{cases}$$
$$\Delta\alpha = \alpha\left(a_i.\text{alarm\_event}\right) - \alpha\left(a_j.\text{alarm\_event}\right). \tag{2}$$

α (a_i.alarm_event) indicates the phase of a_i's attack event, and $\Delta\alpha$ is the difference between the phases of the two alarms. From the attacker's point of view, the attack of subsequent phase is more complex and purposeful and has higher access permission, so if $\Delta\alpha$ equals to 0 or 1, the degree of correlation between these two attack events is higher.

(2) The IP Similarity Function

$$F_{\text{IP}}\left(a_i, a_j\right) = \frac{N}{32}, \tag{3}$$

where $N = \max\{H(a_i.\text{sIP}, a_j.\text{dIP}), H(a_i.\text{sIP}, a_j.\text{sIP}), H(a_i.\text{dIP}, a_j.\text{sIP})\}$, sIP means the source IP, dIP means the destination IP, and $H(\text{IP1}, \text{IP2})$ is maximum same digits of the two IP from the high to low in binary. If two alarms have the same source IP or the same destination IP, or IPs of two alarms are in the same network domain, the two alarms may belong to an attack. Such as, if two alarms have different sIP, but the same dIP, then the attack is launched against the same host, for example, the alarm to an attack with a fake source IP address, such as Syn_flood.

(3) The Timestamp Similarity Function. APT attackers do not tend to profit in a short time, instead, they use the "controlled host" as a stepping stone for persistent searching until a thorough grasp of the target is achieved. In an attacking process, the time interval is relatively short when two attacks are in the same phase, and the time interval may be longer when two attacks occur in different phases, and when there is a long latency following the previous access. For this reason, we do not set time window for alarm logs. The similarity function of the timestamp property is as follows:

$$F_{\text{time}}\left(a_i, a_j\right) = e^{-\Delta t},$$
$$\Delta t = a_i.\text{time} - a_j.\text{time}, \text{ the unit of } \Delta t \text{ is day}. \tag{4}$$

The complete similarity is calculated using the following function:

$$F\left(a_i, a_j\right) = \delta_{\text{alarm}_{\text{event}}} F_{\text{alarm}_{\text{event}}}\left(a_i, a_j\right) + \delta_{\text{IP}} F_{\text{IP}}\left(a_i, a_j\right) + \delta_{\text{time}} F_{\text{time}}\left(a_i, a_j\right). \tag{5}$$

IDS alarm logs are in ascending order by the timestamp, and the similarity of two alarms is calculated using the complete similarity function with multidimensional properties. When the similarity is greater than the threshold value, two alarms are considered triggered by the same attack.

3.3.2. Clustering Algorithm Process

Input: alarm log ALARMS = $\{a_1, a_2, a_3, \ldots, a_n\}$, and attack sequence set ASS = $\emptyset$.

Output: attack sequence set ASS = $\{\text{AS}_1, \text{AS}_2, \ldots, \text{AS}_q\}$, where each attack sequence $\text{AS}_i = \langle a_1, a_2, a_3, \ldots, a_n \rangle$ is a set of alarms and reflects a probable APT attack.

① For each original alarm a_i, calculate its membership to each attack sequence AS_i. If the attack sequence set ASS = $\{\text{AS}_1, \text{AS}_2, \ldots, \text{AS}_q\}$ is empty, then make $\text{AS}_1 = \{a_i\}$, and repeat step ①. If ASS is not empty, then use AS_1 in the ASS set in step ②.

② Scan attack sequence $\text{AS}_i = \langle a_1, a_2, a_3, \ldots, a_k \rangle$. First determine whether the phase of the alarm event a_i is equal to or later than the phase of AS_i (the phase in which the latest timestamp in AS_i occurs). If the answer is yes, go to step ③, and if the answer is no, then go to step ④.

③ Calculate the similarity between a_i and each element in AS_i separately using the similarity function and use the maximum value of the results as a membership degree of a_i to AS_i. If the membership degree is greater than or equal to the preset threshold value λ, then add a_i to attack sequence $\text{AS}_i = \{a_1, a_2, a_3, \ldots, a_k, a_i\}$ and go to step ④.

④ Take the next AS_i in ASS, if it exists, repeat step ②; if not, it means that all the attack sequences in the ASS have been scanned. If the membership degree of a_i to every attack sequence is less than λ, then create a new element $\text{AS}_r = \{a_i\}$ and add AS_r to ASS = $\{\text{AS}_1, \text{AS}_2, \ldots, \text{AS}_q, \text{AS}_r\}$, before going to step ⑤.

⑤ Repeat step ① to step ④ above until all ALARMS are analyzed.

3.4. Mining Out the Attack Scenario. We filter attack sequence set combining purpose and phase characteristic of APT attack and delete the incomplete attack sequence of all IP addresses not involving key assets. In the process of converting an attack sequence into a directed graph, alarms with timestamps approximate to the same attack event are merged into one attack event node. This is because the attacker would use different automation tools during the attack to make continuous malicious requests, generating alarms temporally approximate to the same attack event. Finally, the multiple directed graphs are converted to an attack scenario graph through the probability transfer matrix. The key steps of mining algorithm are shown in Figure 3.

Input: attack sequence set ASS = $\{\text{AS}_1, \text{AS}_2, \ldots, \text{AS}_n\}$ and the IP set IIP of key assets

Output: attack scenario graph

① Get a new $\text{AS}_i = \langle a_1, a_2, a_3, \ldots, a_n \rangle$ from ASS, $a_1, a_2, a_3, \ldots, a_n$ is sorted by timestamp. Determine whether the phase of the last alarm in AS_i is ahead of phase 3, and whether the length of AS_i equals 1. If one of the two conditions is met, and none of the IPs of AS_i is in IIP (key asset IP), then discard this attack sequence and repeat step ①; otherwise go to step ②.

② Convert the first alarm in AS_i to an event node that contains the alarmed attack event and scan from the second alarm, before going to step ③.

③ Take unspecified alarms in turn as a_i, and determine if its corresponding attack event is the same as the attack event of the previous alarm a_j. If the answer is yes then do not create a new node and repeat step ③, or, if the answer is no, convert a_i to a node that contains an attack event, and add a side from a_j node to a_i node. Repeat ③ until the last alarm in AS_i is processed. Then use the matrix to save the directed graph.

④ Go to step ①, if all attack sequences have been analyzed go to step ⑤.

⑤ Initialize a transfer matrix of an empty attack event, where each row and each column represents an attack event in the directed graph. Scan each directed graph, and if there is a directed side between the two attack event nodes A and B, add 1 to the value of location (A, B) in the matrix, and if a new attack event cannot be found in the matrix, add a row and a column in the transfer matrix to represent this attack event. Then each numeric in a row is converted to its proportion to the sum of all the numerical values in that row. The final matrix is expressed in the form of a directed graph.

4. Experiment Analysis

In order to prove the effectiveness of our method for mining out APT attacking scenario, we have used the IDS monitoring environment of a certain company and have simulated 10 advanced persistent attacking processes to steal data. Firstly, we use advanced transverse scan to probe an attack and exploit the vulnerabilities of the key hosts so as to increase access permission. In the process, attack tools such as Nmap, Sqlmap, and chopper are used in sending emails with malicious attachments and exploiting vulnerabilities. For example, Namp is used to scan multiple machines, and some of the vulnerable hosts are targeted with further attacks to extract permission. The entire process lasted one month, during which time IDS alarm logs were collected, and there are 1000 or so valid alarms after false positives were eliminated, some of which have similar timestamps.

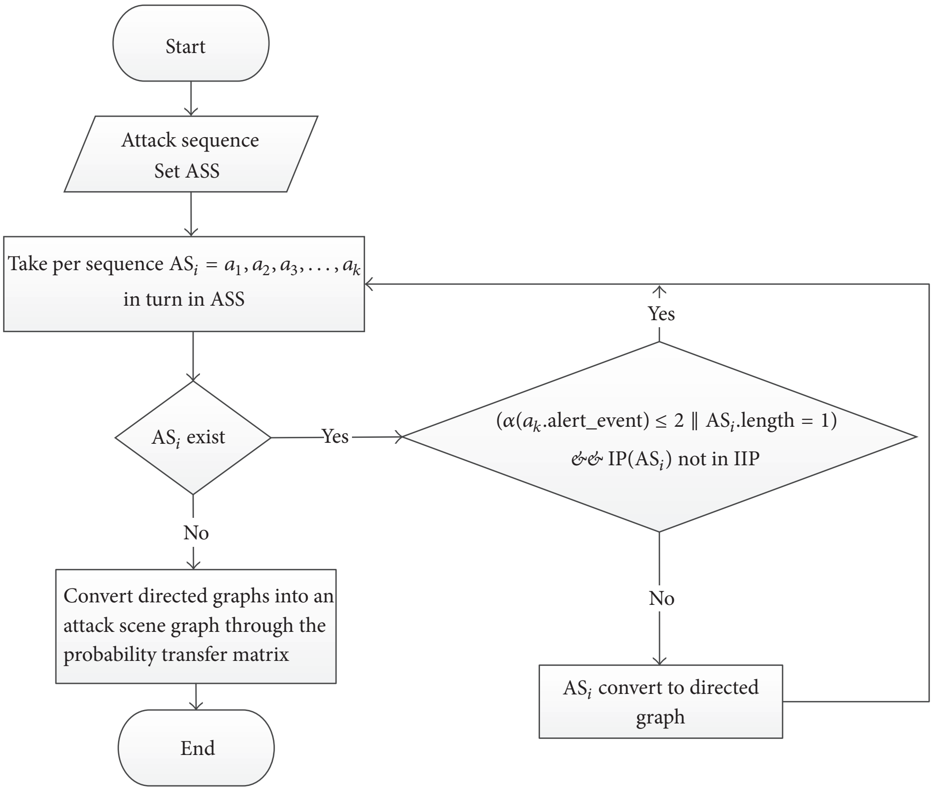

Start
Attack sequence
Set ASS
Take per sequence $AS_i = a_1, a_2, a_3, \ldots, a_k$
in turn in ASS
AS_i exist
Yes
$(\alpha(a_k.\text{alert\_event}) \leq 2 \parallel AS_i.\text{length} = 1)$
&& $IP(AS_i)$ not in IIP
Yes
No
AS_i convert to directed graph
No
Convert directed graphs into an attack scene graph through the probability transfer matrix
End

Figure 3: The process of mining out the attack scenario.

Different alarms to the same attack event are generated by the automated tools.

We then classify all the attack events in the experimental data into four phases of an APT attack before carrying out fuzzy clustering.

4.1. Clustering Algorithm Implementation. Alarms are sorted by timestamp in temporal order. Each alarm is normalized to the six-tuple format as in Definition 1.

The similarity function with multidimensional properties defined in Section 3.3.1 and the clustering algorithm defined in Section 3.3.2 are then used to cluster the IDS alarms ($\lambda = 0.65$, $\delta_{\text{alarm\_event}} = 0.4$, $\delta_{\text{IP}} = 0.4$, $\delta_{\text{time}} = 0.2$).

After clustering, 25 attack sequences are formed, some of which only have scanning and probing behaviors.

4.2. Attack Scenario Mining. We analyze the set of attack sequences generated above and discard any attack sequence where the last alarm is generated in the first or second phase of an attack event. There were 14 relatively complete attacks. Since attackers used different network hosts to launch attacks during the APT attack, attack alarms generated in one attack-planning process were distributed into different attack sequences, resulting in incomplete attack sequences. Eight attack sequences that conformed to the planning process were found after analyses were made. The eight attack sequences were then converted to directed graphs according to the algorithm in Section 3.4. Some directed graphs are shown in Figure 4.

The probability transfer matrix corresponding to all attack sequences is shown in Figure 5.

The probability transfer matrix is then converted into an attack scenario as shown in Figure 6.

Figure 6 shows that we can construct attack scenario by our proposed method. For an attack sequence, the attack means gradually changing from elementary to advanced, obtained permissions are getting more and more powerful, and suspicious files transmission or Trojan back door connection should happen in the end, which meet the phased characteristic of an APT attack. In order to verify the validity of the mined attack scenario, we analyze the attack scenario of an APT attack case named "Sea Lotus" detected by a certain organization. The attack event was unfolding when an intranet host user clicked the malicious mail attachment disguised as a normal file, resulting in the server's terminal virus infection and being controlled by an illegal APT organization, who then implanted the Trojan file qq.exe.bak in the folder c:\users\user\appdata\roaming\tencent, where communication was made with a hacker IP address and a small-amount data transmission was done. By analysis of the whole APT process, we find a series of events including alarms against a large number of malicious mail attachments, DNS requests from malicious domains, suspicious file transfers, and malicious domain connections. These events

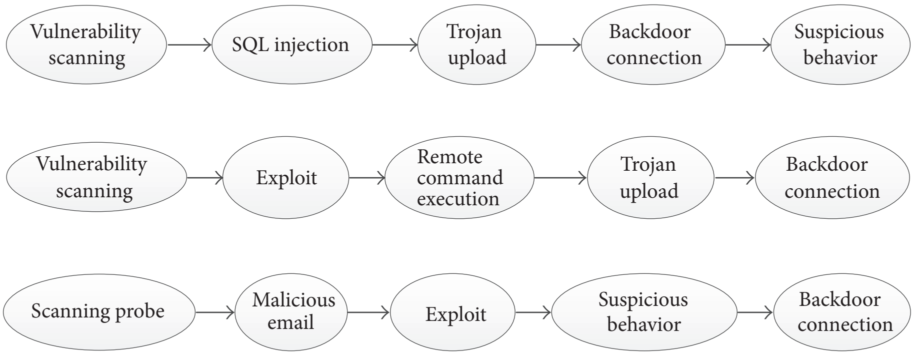


Figure 4: Part of the directed graph of an attack sequence.

	Scanning probe	CGI attack	Trojan upload	Exploit	Backdoor connection	Suspicious behavior	Remote command execution	Malicious email	Suspicious file directory access
Scanning probe		0.25		0.5				0.25	
CGI attack			0.33			0.67			
Trojan upload					0.5		0.2		0.3
Exploit						0.33	0.33		0.33
Backdoor connection						1			
Suspicious behavior					1				
Remote command execution			0.33						0.67
Malicious email				0.6		0.4			
Suspicious file directory access					1				

Figure 5: The probability transfer matrix.

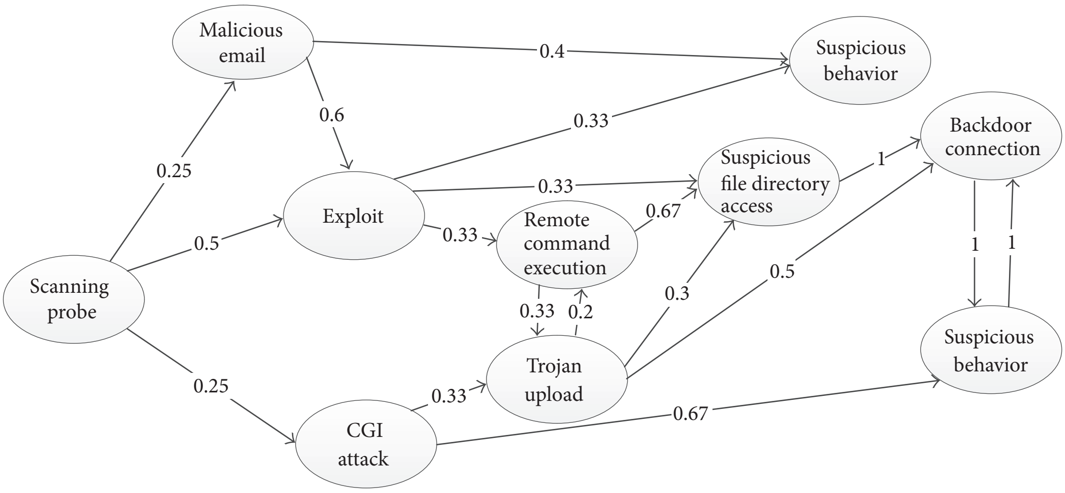


Figure 6: An attack scenario.

conform to the mined APT attack scenario, meaning that the attack scenario we mined reflects the true APT attack chain and is useful for the detection and defense of an APT attack.

Additionally, we use accuracy rate $R_r = N_c/N_n$ to evaluate the APT attack scenario mining method, where N_c is the effectively mined attack sequence by our mining method and N_n is the APT attack sequence that should be mined out. All the mined attack sequences include some attack sequences that do not match our attack strategy, and we delete such attack sequences, R_r= 80%.

Feng et al.'s paper [11] used alert clustering based on the correlation of IP addresses to produce alarm cluster sets. We

Table 2: Results of clustering algorithms with different dimensions.

Clustering algorithm	Dimensions of clustering algorithm	Result
Fuzzy Clustering which includes the attack-event dimension	alarm_event, IP address, timestamp	Escalating attack mechanism, and there is higher correlation in an attack sequence
Fuzzy Clustering which exclude the attack-event dimension	IP address, timestamp	Attack events intersect, and there is small correlation in an attack sequence

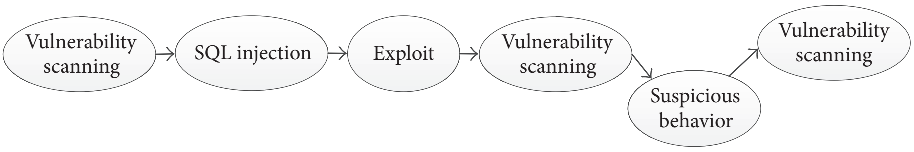


Figure 7: Attack sequence fragment.

use the clustering algorithm excluding the attack event dimension to process the same experimental data and do not classify attack events. We cluster with IP address and timestamp and analyze attack sequence set without considering attack events. We can get an attack sequence as shown in Figure 7.

In Figure 7, detection scanning occurs after either vulnerability exploitation or suspicious behaviors and attack events of different phases intersect. The clustering method that uses only two dimensions of the IP and the timestamp tends to correlate the attacking processes on the same asset by multiple attackers and/or certain misoperations to one attack sequence, resulting in decreased correlation between different alarms in an attack sequence. The results of clustering algorithms with different dimensions are shown in the Table 2.

By adding an extra dimension of the attack event, our proposed method can reduce the occurrence of decreased correlation. Thus, our method increases the degree of correlation between different alarms in an attack sequence, and it does not rely on any attack model built with expert knowledge.

5. Conclusion

In this paper, the attack events in an IDS log are classified based on the IKC model, the method of fuzzy clustering is used to correlate the alarm logs to produce the attack sequence set, and the nonaftereffect property of the probability transfer matrix is used to excavate the attack scenario from the attack sequence set. Based on the phased characteristic of an APT attack, in this paper, the purpose of an APT attack in each phase is analyzed and attack events are classified. In addition to the IP address and the timestamp, the use of the attack event as another key dimension in fuzzy clustering also improves the correlation degree of alarms in the same attack sequence. The effectiveness of this method has been proved by experiments. The method proposed in this paper can automatically construct attack scenario based on IDS logs and the attack scenario provides guidance for the detection and defense of APT attacks.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National key Research and Development Program of China (2016YFB0800903) and the NSF of China (U1636212, U1636112).

References

[1] P. Chen, L. Desmet, and C. Huygens, “A study on advanced persistent threats,” in *Communications and Multimedia Security: 15th IFIP TC 6/TC 11 International Conference, CMS 2014, Aveiro, Portugal, September 25-26, 2014. Proceedings*, vol. 8735 of *Lecture Notes in Computer Science*, pp. 63–72, Springer, Berlin, Germany, 2014.

[2] I. Friedberg, F. Skopik, G. Settanni, and R. Fiedler, “Combating advanced persistent threats: from network event correlation to incident detection,” *Computers & Security*, vol. 48, no. 7, pp. 35–57, 2015.

[3] X.-S. Zhang, W.-N. Niu, G.-W. Yang et al., “Method for APT prediction based on tree structure,” *Journal of University of Electronic Science and Technology of China*, vol. 45, no. 4, pp. 582–588, 2016.

[4] SkyEye: OceanLotus APT Report [2015-05-29], https://ti.360.net/static/upload/report/file/OceanLotusReport.pdf.

[5] K. Munro, “Deconstructing flame: The limitations of traditional defences,” *Computer Fraud and Security*, vol. 2012, no. 10, pp. 8–11, 2012.

[6] C. Choi, J. Choi, and P. Kim, “Abnormal behavior pattern mining for unknown threat detection,” *Computer Systems Science & Engineering*, vol. 32, no. 2, pp. 171–177, 2017.

[7] Y. Fu, H. LI, X.-p. Wu, and J. Wang, “Detecting APT attacks: a survey from the perspective of big data analysis,” *Journal on Communications*, vol. 36, no. 11, pp. 1–14, 2015.

[8] C. Tankard, “Advanced Persistent threats and how to monitor and deter them,” *Network Security*, vol. 2011, no. 8, pp. 16–19, 2011.

[9] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, “Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains,” in *Proceedings of the 6th International Conference on Information Warfare and Security (ICIW ’11)*, pp. 113–125, Curran Associates Inc, Washington, Wash, USA, March 2011.

[10] H.-B. Mei, J. Gong, and M.-H. Zhang, “Research on discovering multi-step attack patterns based on clustering IDS alert sequences,” *Journal on Communications*, vol. 32, no. 5, pp. 63–69, 2011.

[11] X. Feng, D. Wang, M. Huang, and J. Li, “A mining approach for causal knowledge in alert correlating based on the markov property,” *Jisuanji Yanjiu yu Fazhan/Computer Research and Development*, vol. 51, no. 11, pp. 2493–2504, 2014.

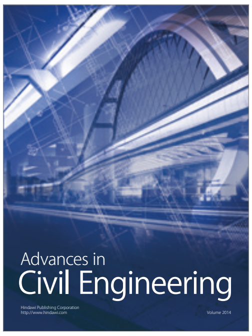
Journal of
Engineering

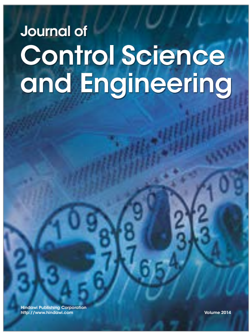
The Scientific World Journal

International Journal of
Rotating Machinery

Journal of
Sensors

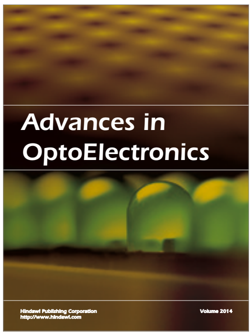
International Journal of
Distributed Sensor Networks


Advances in
Civil Engineering


Journal of
Control Science and Engineering

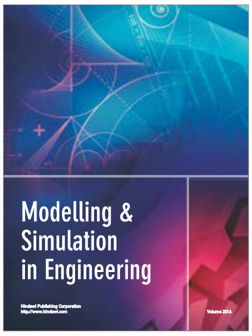
Journal of
Robotics

Journal of
Electrical and Computer Engineering

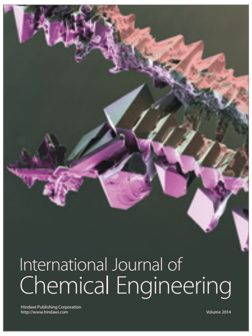

Advances in
OptoElectronics

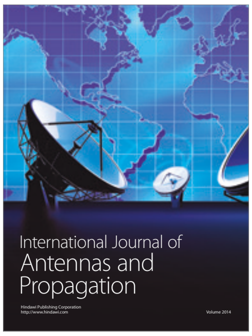
VLSI Design

International Journal of
Navigation and Observation

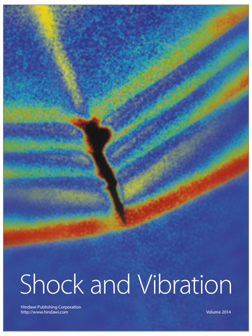

Modelling & Simulation in Engineering

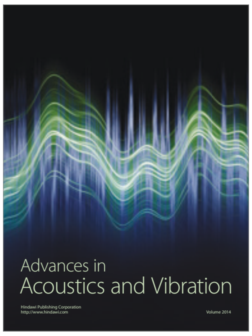
International Journal of
Aerospace Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and Propagation

Active and Passive Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration