A Security Event Correlation Algorithm Based On Attack Sequence

Zhang Dedong[1], Wang Hongwei[1],Feng Kailiang[1]

[1]Institute of Computing Technology, China Academy of Railway Sciences, Beijing 100081, China

[a]email: zhdd0411@163.com

Keywords: Security event, Attack sequence, Association analysis, Security operation center

Abstract. A new multi-stage attack correlation method based on attack sequence is proposed in this paper. The algorithm first excavates the attack sequence of network attack behaviors from a large number of security events, and then analyzes the correlation of the events which are in accord with certain attack pattern using membership function. The simulation results show that the algorithm can not only correlate multiple isolated security events in attack scenarios to detect composite attack, but also can find the real security threat hidden in security events.

Introduction

With the continuous development of network technology, malicious attacks, data reveal, computer viruses and other network security threats emerge in endlessly. In response to the network security threats, the enterprise has deployed the firewall, intrusion detection system, vulnerability scanning system and other security equipment. However, most of these devices have specific function and do not have an effective working relationship. More seriously, the devices can produce huge amounts of the alarm information in operation. It is difficult to find the real security events from the huge information. How to analyze these security events and maximize these devices' effect?. Therefore, the security operation center (SOC) came into being. In the SOC, the security event correlation analysis is the key technology. The technology can find the relationship between events and the real events. The research of correlation analysis has attracted widespread attention.

In 2001, Debar and Wespi[1] proposed a correlation analysis algorithm based on rules for intrusion detection alerts, and the algorithm was applied in the IBM Tivoli system. In 2002, Peng Ning[2] from the university of north Carolina proposed a correlation algorithm by constructing attack scenarios to define the correlation rule. Literature [3] proposed a correlation analysis method based on rules, which implemented the content security event correlation analysis. However, these correlation algorithms based on the rules has some defect. Such as: the correlation rules are complex, require professionals to develop and are not suitable for the unknown attacks. In 2002, IBM Zurich laboratory [4] put forward a security incident aggregation method based on attribute similarity. This method could reduce the alarm rate of false positives, but were prone to alarm omission. Wang Lingyu et al[5] proposed a correlation algorithm based on attack graphs. The security events were mapped to the attack graphs in chronological order, and then dig event correlation by using the graph theory knowledge in the algorithm. Roschke et al[6] proposed A new multi-stage attack correlation method by finding the shortest path in the the attack graph. Kuang Qingyuan et al[7] proposed security event correlation algorithm based on attack intention. Ourston et al[8]proposed security event correlation algorithm based on the hidden markov models. Liu jing et al[9] proposed a security events analyze algorithm method based on neural networks and genetic. These algorithms [4-9] were based on machine learning theory, which need a larger of training data. However, due to the diversity of the means of attack, training data did not necessarily apply to the real network environment and resulted in larger non-response rates or the rate of false positives. Liu Lan et al[10] proposed a causality correlation algorithm based on fuzzy equivalent restriction, and it could find potential security threats in a large number of incomplete and fuzzy events. However, this algorithm had high time complexity. In view of this, a new multi-stage attack correlation method based on attack sequence is proposed in this paper. The algorithm first excavates the attack sequence of network attack behaviors from a large number of security events, and then analyzes the

correlation of the events which are in accord with certain attack pattern using membership function. The simulation results show that the algorithm can not only correlate multiple isolated security events in attack scenarios to detect composite attack, but also can find the real security threat hidden in security events.

Multi-stage Attack Analysis

Multi-stage attack instances.(1) Worm attack--Sasser. Sasser was caused by buffer overflow vulnerabilities of Laass service. The attacker could control system permissions by the vulnerabilities. The specific attack steps are as follows.

Step 1: The attackers scan Windows system, to find whether a the Lsass service port (445) is open or not;

Step 2: If the port is open, the attackers launch an attack to the target host by the Sasser virus and infect the host;

Step 3: the Sasser virus driver the host download a file called "_up.exe" files", so as to control system permissions.

(2) Infiltration attack of obtaining the manager access. The attack steps are as follows.

Step 1: %cp/bin/csh/usr/spool/mail/root, copy the /bin/csh/ shell to usr/spool/mail/root;

Step 2: % chmod 4755 /usr/spool/mail/root, change the permissions of usr/spool/mail/root;

Step 3: %touch x, create an empty file;

Step 4: %mail root < x, send an empty file to the root directory

Step 5: % /usr/spool/mail/root, execute setuid-to-root shell.

The characteristics of multi-stage attack. Multi-stage attack has the following features

(1) The attacker has a strong purpose, and use a variety of methods of attack. For the same attack, the attacker might use a variety of methods, such as: the method of using SQL injection, the method by using the vulnerabilities of port, the method by using the vulnerabilities of buffer overflow, the method by using the vulnerabilities of protocol.

(2) The attack event can be divided into some steps. The previous steps are always preparing for the steps behind, until reach the final purpose.

(3) There are some temporal relations between the attack steps. Attack steps carried out in accordance with the time sequence, and the beginning of a step often ended a step above the premise. There is certain timeliness between steps and an attack result may not work beyond a certain time interval.

A Security Event Correlation Algorithm Based on Attack Sequence

Related knowledge. The events are generated by the device called the original events. The original events after formatting, polymerization, merging become senior security events. In this paper, all the events refer to senior security events.

Definition 2.1. Attack sequence: a set of events with chronological order. Let $es =< e_1,...,e_i,...,e_l > (1 < i < l)$ stand for attack sequence, where, e_i refers to a event.

Definition 2.2. Sequence inclusion: for any two attack sequences $es_1 =< e_1,..,e_i,..,e_n >$ and $es_2 =< f_1,.., f_i,.., f_m >$, attack sequence es_1 contains es_2 if and only if meets the following conditions, (1) $m \le n$; (2) $e_1 = f_1, e_{i_2} = f_2, e_{i_3} = f_3,..., e_{i_n} = f_m \quad (i_2 < i_3 < ... < i_n)$.

Definition 2.3. Global attack sequence: a set of all the events with chronological order.

Definition 2.4. Duration of attack sequence: let $es =< e_1,...,e_i,...,e_l > \quad (1 < i < l)$ stand for a attack sequence. e_1 is the first attack event in the sequence. e_l is the last attack event in the sequence. The duration of attack sequence can be expressed as $W_T = e_l.end\_time - e_1.begin\_time$,

Definition 2.5. The candidate of attack sequence: Divide the global attack sequence into some attack sequence with chronological order.

Definition 2.6. The support of attack sequence: if the attack sequence es_1 contains es_2, it is said

es_1 support es_2. The support of attack sequence: the number of supported candidate attack sequences proportion of all candidate attack sequences

Definition 2.7. Maximal attack sequence: the sequence is not be contained in other sequence.

Definition 2.8. Fuzzy causal relation: for any security event e_i, let $A_i = \{\mu_1, \mu_2, ..., \mu_n\}$ stand for security attribute of e_i. If the event time interval within time of duration, and existing $e_1.\mu_1 = e_2.\nu_1 \wedge e_1.\mu_2 = e_2.\nu_2 \wedge ... \wedge e_1.\mu_k = e_2.\nu_k \quad (k \le n, k \le m)$, it is said, e_1 and e_2 have fuzzy causal relation.

Definition 2.9. Membership functions: let membership functions $\sigma(e_1, e_2)$ describe the causal relationship between security events. Membership functions $\sigma(e_1, e_2)$ can be described as follows:

$$\sigma(e_1, e_2) = \frac{\prod_{i=1}^{q} \omega(\mu_i, \nu_i) \times \sum_{j=1}^{k} \omega(\mu_j, \nu_j)}{\sum_{i=1}^{Mat(e_1, e_2)} w_i} (0 < q \le k \le \min(n, m)) \quad (3\text{-}1)$$

Where μ_i is the security attribute of e_1 and ν_i is the security attribute of e_2. $Mat(e_1, e_2)$ refers to the security attribute number of the e_1 and e_2 with matched. w_i refers to attribute weights. $\omega(\mu_i, \nu_i)$ can be calculated as follows:

$$\omega(\mu_i, \nu_i) = \begin{cases} 1 & \mu_i = \nu_i, \mu_i \in B, \nu_i \in B \\ w_i & \mu_i = \nu_i, \mu_i \notin B, \nu_i \notin B \\ 0 & \mu_i \ne \nu_i \end{cases} \quad (3\text{-}2)$$

Algorithm description. The algorithm based on attack sequence is divided into four steps, namely: construct global attack sequence, generate candidate attack sequence, calculate the maximal attack sequence, calculate security event causality.

Step 1: Construct global attack sequence. Let D stands for the set of security events. The events can be sorted according to the time order and labeled according to event type. The global attack sequence can be defined all the events sorted according to the time order. The following instances constitute a attack sequence. The global attack sequence can be described δ =(4,7,11,13,8,5,7,4,11, 13,9,10, 14,9 ,15,2).

Table 1 attack instances

Time	Type	Other attributes	label
2016-3-15 09:24:29	Port scanning	……	4
2016-3-15 11:02:37	Configuration changes	……	7
2016-3-16 04:22:01	Trojan invasion	……	11
2016-3-17 01:42:29	information stealing	……	13
2016-3-17 08:29:12	Website Distortion	……	8
2016-3-17 12:03:31	authentication failure	……	5
2016-3-17 23:22:09	Configuration changes	……	7
2016-3-18 08:12:33	Port scanning	……	4
2016-3-18 13:24:10	Trojan nvasion	……	11
2016-3-18 18:14:54	information stealing	……	13
2016-3-18 21:10:15	Equipment Failure	……	9
2016-3-19 07:12:34	Virus attacks	……	10
2016-3-19 11:27:51	InformationMasquerading	……	14
2016-3-19 15:14:55	Equipment Failure	……	9
2016-3-19 21:10:16	Information Masqueradin	……	15
2016-3-20 05:12:35	worm virus	……	2

Step 2 Generate candidate attack sequence. The global attack sequence can be divided into some candidate attack sequence according to the duration W_T. It is described $\delta = (\delta_1, \delta_2, ..., \delta_n)$, and δ_i is a candidate attack sequence. Where $\delta_i = < e_i, e_{i+1}, e_{i+2}, ..., e_{i+m} > \quad (1 \le i \le n)$. The duration can be

described in formula (3-3).

$$(e_{i+m}.end\_time - e_i.begin\_time \le W_T) \wedge (e_{i+m+1}.end\_time - e_i.begin\_time > W_T) \quad (3\text{-}3)$$

In above instances, the candidate attack sequence might be generated as following.

Table2 the candidate attack sequence of the above instance

NO	The candidate attack sequence
1	<4,7,11,13,8,5>
2	<7,11,13,8,5,7>
3	<11,13,8,5,7,4>
4	<13,8,5,7,4,11,13>
5	<8,5,7,4,11,13>
6	<5,7,4,11,13,9>
7	<7,4,11,13,9,10>
8	<4,11,13,9,10,14,9>
9	<11,13,9,10,14,9>
10	<13,9,10,14,9,15>
11	<9,10,14,9,15,2>

We can calculate all the attack sequence that satisfies the minimum support from the candidate attack sequence according to the minimum support and definition 2.6. In above instances, suppose that the minimum support is set 15%, the attack sequence that satisfy the minimum support can be described in table 3.

Table 3 The attack sequence satisfying the minimum support

attack sequence	support
<4>	2/11
<4,11>	2/11
<4,13>	2/11
<7,11>	2/11
<7,13>	2/11
<11,13>	2/11
<4,11,13>	2/11
<7,11,13>	2/11

Step 3 Calculate the maximal attack sequence. We can calculate the maximal attack sequence according to the definition 2.7. In above instances, the maximal attack sequences are <4,11,13> and <7,11,13>.

Step 4 Calculate security event causality. Event type can not only determine an attack event. If we want to determine the causality between the events, we should calculate the causality of all the attribute synthetically. Let T_σ stands for the membership threshold, e_i stands for a security event in the maximal attack sequence. The causality between the events e_i and e_j can be calculated according to membership functions $\sigma(e_i, e_j)$ (formula 3-1 and formula 3-2). If the membership functions $\sigma(e_i, e_j) > T_\sigma$, it is shows that the events(e_i, e_j) have causality.

Analysis of Experimental Results

Experimental environment. In order to verify the validity of the algorithm, the test experiment is set up to verify the algorithm. The experimental environment is shown in figure 1,where the host(172.22.51.47) is attack client, the host (172.16.18.39，172.16.18.40,172.16.18.41，172.16.18.42)

is puppet machine, the host (172.20.96.55)is the victim host and IDS is used to record security events.

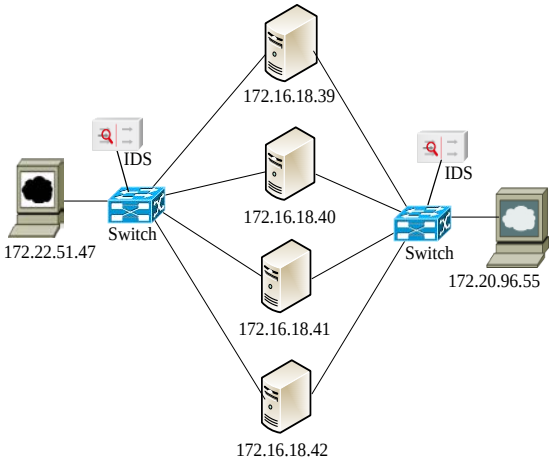

Fig.1. Experimental Topology

Test data uses the DARPA data set, which is from MIT Lincoln lab. Data set LLDOS 1.0 contains a complete attack sequence and the attack contains 5 steps. The experiment process is as follows:(1)The attacker uses the IPSweep script to scan puppet machine(172.16.18.39，172.16.18.40，172.16.18.41，172.16.18.42) for collecting information; (2)The attacker launch a sadmid attack to puppet machine from attack client; (3)The attacker launch a sadmind buffer overflow attack to puppet machine; (4)The attack installs the host file and DDOS Daemon to puppet machine by telnet; (5)Puppet machine launch stream DDOS attack to victim host; (6)Attack for 20 seconds, the host breaks down.

Results analysis. Set the support of attack sequence 15%, membership threshold 0.2, the candidate of attack sequence 2s. In the process of experiment, there are 4216 original events. After formatting, and merging, there are 288 senior events and 262 candidate attack sequences. There 94 attack sequences contained in the candidate attack sequences, the senior events contains some events of IPSweep script scanning, sadmind attack, buffer overflow attack, telnet remote operation and mstream DDOS attacks. The support of attack sequence is 36%. The above experimental results show that the algorithm can extract security attack sequence in reasonable support and threshold.

Conclusion

Security events are the important factor of security situational awareness and risk assessment. Correlation analysis is an important means of mining the real security events. A new multi-stage attack correlation method based on attack sequence is proposed in this paper. The algorithm first excavates the attack sequence of network attack behaviors from a large number of security events, and then analyzes the correlation of the events which are in accord with certain attack pattern using membership function. The simulation results show that the algorithm can not only correlate multiple isolated security events in attack scenarios to detect composite attack, but also can find the real security threat hidden in security events.

Reference

[1] Debar H, Wespi, A. Aggregation and Correlation of Intrusion-Detection Alerts [A]. Recent Advances in Intrusion Detection, International Symposium [C], Ca, Usa, October 10-12, 2001, Proceedings. 2001.pp. 50-78.

[2] Ning P, Cui Y, Reeves D S. Constructing attack scenarios through correlation of intrusion alerts[A]. ACM Conference on Computer and Communications Security [C], 2002, pp.245-254.

[3] Ge Lin, Ji Xinsheng, Jiang Tao. Discovery of network information content security incidents

based on association rules and its implementation in map-reduce [J]. Journal of Electronics & Information Technology, 2014, 36(8):1831-1837.

[4] Julisch K. Clustering intrusion detection alarms to support root cause analysis [J]. Acm Transactions on Information & System Security, 2003, 6(4):443-471.

[5] Wang L, Liu A, Jajodia S. Using attack graphs for correlating, hypothesizing, and predicting intrusion alerts [J]. Computer Communications, 2006, 29(15):2917-2933.

[6] Roschke S, Cheng F, Meinel C. A new alert correlation algorithm based on attack graph [A]. Computational Intelligence in Security for Information Systems International Conference [C], Iwann, Proceedings. 2011.pp.58-67.

[7] Kuang Qingyuan, Wu Bin, Wu Chunhua. A security event correlation algorithm based on attack intention[A]. The 19th National Youth Academic Essays of Communication [C]. 2015.pp.248-253.

[8] Harahap E, Sakamoto W, Nishi H. Failure prediction method for network management system by using bayesian network and shared database[A]. Information and Telecommunication Technologies 2010 8th Asia-Pacific Symposium on IEEE [C]. 2010.pp.1-6.

[9] Liu Jing, Gu Lize, Niu Xinxin. Network security events analyze method based on neural networks and genetic algorithm [J]. Journal of Beijing University of Posts and Telecommunications, 2015, 38(2):50-54.

[10] Liu Lan, Wen WuShao, Xu Xiaoping. On causality correlation algorithm based on fuzzy equivalent restriction of security event [J]. Computer Applications and Software, 2011, 28(3): 133-136.