Attack Intent Analysis Method Based on Attack Path Graph

Shijin Li
College of Mathematics and Computer Science,Fuzhou University
Fuzhou,Fujian,China
Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing
Fuzhou,Fujian,China
+8618120802689
18120802689@163.com

Minchen Zhu
College of Mathematics and Computer Science,Fuzhou University
Fuzhou,Fujian,China
zmc@fzu.edu.cn

Yanbin Qiu
College of Mathematics and Computer Science,Fuzhou University
Fuzhou,Fujian,China
Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing
Fuzhou,Fujian,China
+8618659758006
314425969@qq.com

ABSTRACT

At present, with the increase of automated attack tools and the development of the underground industrial chain brought by network attack, even well-managed network is vulnerable to complex multi-step network attack, which combines multiple network vulnerabilities and uses the causal relationship between them to achieve the attack target. The detection of such attack intention is very difficult. Therefore, in order to solve the problem that the real attack intention of the attackers in complex network is difficult to be recognized, this paper proposes to assume the possible targets in the network according to the important asset information in the network. By constructing the hierarchical attack path graph, the probability of each hypothetical attack intention target is calculated, and the real attack intention and the most likely attack path of the attacker are deduced. The hierarchical attack path graph we use can effectively overcome the cognitive difficulties caused by network complexity and large scale, and can quantitatively and qualitatively analyze the network status. It is of great importance to make the protection and strategy of network security.

CCS Concepts

• **Security and privacy→Network security→Web protocol security**

Keywords

Attack path graph Attack intention Critical assets

1. INTRODUCTION

With the advent of the information age, the scale and application of the network continue to expand, and the malicious attacks of network information systems become more diverse and complicated. At the same time, massive alarm information and complex network scale make network analysis very difficult. These have brought enormous challenges to the attacker's identification of real intentions and network security protection.

ICCNS 2018, November 2–4, 2018, Qingdao, China

ACM ISBN 978-1-4503-6567-3/18/11…$15.00
https://doi.org/10.1145/3290480.3290502

In general, attackers are rational, do not blindly launch attacks, but a series of strategies and actions to achieve an intention. If you can find the logical relationship between the attack chains from a large number of attacks [1], build an attack graph to measure the overall security of the network [2]. Inferring the attacker's true attack intention, it can greatly improve the network security protection ability. And provide support and guidance for understanding network security posture, planning and designing network protection strategies.

The cyberattack intention is heavily dependent on the environment. In a specific network environment, the attack intentions that can be achieved are certain and clear. This paper first proposes a set of intent hypotheses for the security requirements of specific network environments, and constructs a hierarchical attack path map through a causal relationship model between vulnerabilities. The attack path map is used to calculate the probability of the hypothetical attack intention target, and the inference is calculated to calculate the attacker's true attack intention and the most likely attack path.

The main contributions of this paper are as follows: an attack graph model with high efficiency and good scalability is proposed, and it can be applied to large and complex network environment, and is applied to attack intention analysis, and has achieved good results.

2. RELATED WORKS

Bratman[3] first proposed the concept of intention, which pointed out that the role of intention is to guide rational decision-making and plan future behavior. The process of intent recognition is described in [4], which is inferred by observing the behavior of agents and their impact on the environment. From the perspective of defenders, [5] considers the impact of attack behavior on network security protection objectives, classifies attack intentions, and proposes a famous CIA security classification model. Attack intent behavior is classified by breaking the confidentiality, integrity, and availability of the system.

Since the attack graph was proposed and used for network security analysis, the research work mainly focused on attack graph automation, visual construction [6][7] and formal analysis of attack graph. According to the structure of the attack graph, the early attack graph is biased based on state construction in terms of attack graph construction and rendering. The evolution of the global state of the network is described by the state transition caused by the attack behavior represented by the edge, but this method is not conducive to the network-scale system analysis due to the impact of the state space explosion. In-depth study of the

path reachability problem in the attack graph, using the analysis of path reachability, reducing the scale of the attack graph. However, there are two main limitations to the attack graph research: first, the explosion of the attack graph state space caused by large-scale complex networks; second, the reliance on expert knowledge base support.

At present, the representative of the network attack intention analysis method is representative of the event correlation analysis method based on the attack premise and result [8][9]; Based on the Idea of On-line Diagnostics Method [10]; Network attack intent recognition method based on similarity analysis [11]; Real-time attack intent recognition based on fuzzy minimum-maximum neural network [12]; Attack intent recognition method based on multiple integration technologies [13]; The above methods have generally achieved good results, but the resume of the prior knowledge base and the selection of a large number of training samples are more expensive. Inspired by the above methods, we propose an attack intent analysis method based on the attack path graph.

3. GENERATION OF THE ATTACK INTENT ASSUMPTIONS BASED ON CRITICAL ASSETS

Critical asset is an entity that the security administrator focuses on protecting based on the importance of value, including information, services, and equipment. Criticalassets are often the main targets of attackers.This set represents all possible attack intent targets of the attacker. We can formulate corresponding protection strategies in advance according to different attack intentions. The specific method is as follows:

Critical asset collection:ASSET, Security requirement set: SEREQU Network attack intent set:INTENT, Purpose set of attack intention:AIM. Attack intention point set: TARGET.And SEREQU={confidence,integrity,availability},
AIM={ aim_1 , $aim_2, aim_3, aim_4, aim_5$ }, Represents illegal elevation of authority, disclosure of information, tampering with information, denial of service, and illegal use of resources. TARGET={ $target_1, target_2, target_3$, $target_4, target_5$ }, Representing accounts, files, processes, computers, and networks. For $\forall$ intent $\in$ INTENT, there is intent(aim, target). aim $\in$ AIM, target $\in$ TARGET. the illegal promotion of authority is the most intensive attack intention, which not only directly affects confidentiality and integrity, but also may indirectly affect system availability. The mapping relationship between the security requirement attribute and the destination attribute of the attack intention is shown in Table 1:

Table 1 Attribute mapping relationship between security requirement attributes and attack intentions

confident	tegrity	availability
aim_1	aim_1	aim_1
aim_2	aim_3	aim_4
aim_5		

The attack intention hypothesis generation algorithm is divided into the following steps:

(1) In the critical asset collection ASSET, according to its security requirements and Table 1, determine the attribute set AIM_i,for the attack intention of the critical asset $asset_i$.

AIM_i={$aim_{i1}, aim_{i2} \ldots aim_{im}$};

(2) According to the correspondence between the attack intention application point and the critical asset, the attack intention application point attribute of the critical asset is determined, = {$target_{i1}, target_{i2}, \cdots target_{il}$};

(3) According to the number of attacks on the history of critical assets in the IPS log, $RECOURD_i$, $RECOURD_i = \{recourd_{i1}, recourd_{i2}, \cdots recourd_{im}\}$

(4)The elements in the collection AIM and the collection $TARGET_i$ are arbitrarily combined, and the attack intention elements that conflict between the $TARGET_i$ attribute and the operational point attribute are removed, as well as the intentions that are inconsistent with the security requirements.

(5) Repeat the above steps for all assets, and finally get a hypothetical set of attack intentions of the network.

The specific generation process is shown in Figure 1:

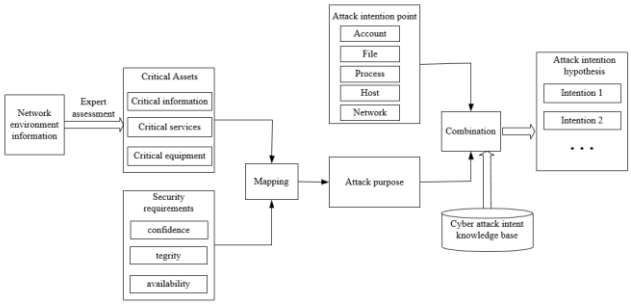

Figure 1 Attack intent hypothesis set generation method

4. FORMATION OF HIERARCHIACAL ATTACK PATH GRAPH BASED ON CAUSAL ANALYSIS

4.1 Vulnerability Causal Analysis Model

By the well-known AVI failure model[5], the attack is a successful use of the system's vulnerability behavior, the purpose is ultimately to cause the entire system's security attributes to fail. Therefore, the vulnerability of the system and the exploitability of the attacker are two necessary conditions for the occurrence of the attack. Moreover, in the process of attack, there is not only a dependency between attack behavior and vulnerability, but also a correlation between vulnerability. Attack is a process of utilizing the next vulnerability on the basis of improving the authority and other benefits obtained by the successful utilization of vulnerability.

We use a simple and efficient vulnerability causal analysis model that can handle large networks of tens of thousands of hosts. The model is based primarily on the preconditions and consequences of the successful use of vulnerability, as shown in Table 2:

Table 2 Vulnerability Premise and Consequence Set

Vulnerability	Attributes	Attribute meaning
Precondition Vuln_pre	Local	Launch a vulnerability attack from the local
	Remote	Launch a vulnerability attack from the network
Consequence	Root	Get administrator or root level permissions
	User	Get guest or user level

Vuln_pos		permissions
	Dos	Launch a denial of service attack
	Other	Other damage to security attributes such as system confidentiality or integrity

The only prerequisite in this model is local or network, and the consequences are only four. Compared with other more complex network attack models, the information needed in this simplified classification method of the model can be obtained from existing Automatic extraction of data, The preconditions are mainly to determine the location of the attacker, whether it is local or in the network. The consequences are based on the rights acquired by the attacker after exploiting the vulnerability (Adimistrator, user, other, Dos) and the textual description caused by the exploit. Such as: "execute arbitrary code", "get system permissions", etc. to determine. This information can be extracted from the existing vulnerability database.

4.2 Attack Path Graph Generation

The hierarchical attack path map will be divided into three levels: the vulnerability layer, the host layer, and the management domain layer.

In the attack path graph generation model, the input initial information is network topology information, vulnerability information, and attack intention hypothesis information generated based on critical assets. The part generation process is shown in Figure 2.

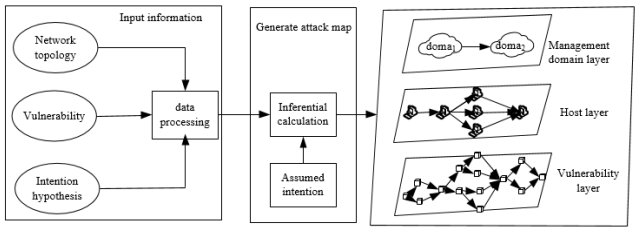

Figure 2 Generation of attack graph

The algorithm for generating the attack path map is as follows:

Input: host set H, host h vulnerability set V_h, host connection relationship C.

Output: attack path graph G(N，E).

(1) In the vulnerability set n of host h, starting from the permission none, search all the $vuln_{i+1} \cdot vuln\_pre \subseteq vuln_i \cdot vuln\_pos$ vulnerability pairs $(vuln_i, vuln_{i+1})$ that satisfy, add $vuln_i$ and $vuln_{i+1}$ to the vertex set, and add edge $vuln_i \rightarrow vuln_{i+1}$ to the edge set until user permission or root permission is reached, that is, generate the attack path graph of host h with none as the initial node and uesr or root as the final node, in which $vuln_i$, $vuln_{i+1} \in V_h$.

(2) It is assumed that starting from the permission of h_0, it satisfies $(h_0, h_1, \theta) \in C$ and can obtain the permission of h_1 through the host . Then h_1 is the new node, edge $h_0 \rightarrow h_1$. It's a new directed edge, $h_1 \in H$.

(3) Suppose D_1,D_2 is two management domains and the attacker obtains the permission of host h_1. Host $h_1 \in D_1$, Host $h_2 \in D_2$, If the attacker can get the permission of h_2, D_2 is the new node, and the edge $D_1 \rightarrow D_2$ is the new directed edge.

(4) In the weak point set, host set and protection domain set of the host respectively, use the breadth-first forward search algorithm to search all the nodes and edges satisfying the condition (1)(2)(3) and add to the corresponding node set N and edge. In set E, the attack path maps of vulnerability level, host level and management domain level are generated.

4.3 Attack Intent Accessibility Analysis Model

Accessibility is an important prerequisite for building an attack graph. It refers to which ports of all the hosts of all the hosts in the network can establish TCP or UDP connections for communication. This involves complex firewall rules and network topologies, and even the accessibility between many two hosts is more computationally intensive than the generated attack path graph. This paper proposes an improved accessibility calculation model.

A reachability matrix is constructed with all ports between each host, where each row in the matrix represents a source port number and each column represents a destination port number. Each cell of the matrix contains a Boolean value indicating whether there is reachability between the source port and the destination port. The specific process has the following steps:

(1) Through the network topology and firewall rule cross-reference verification method, traverse each port in the matrix to determine whether it is connectable.

(2) Calculate the reachability from the source port (the rows in the matrix), calculate only the hosts that the attacker can actually attack at the user or administrator level, and remove some hosts that cannot be attacked from the attack graph. This method can remove a large number of redundant paths and save a lot of computing resources and time when constructing the attack graph.

(3) Grouping hosts with similar reachability in the matrix, and ports that need to pass through firewall rules in different subnets can be classified as filtering reachability domains. For each set of interfaces in the filtered reachability domain, reachability needs to be calculated only for one interface of the set, and all other members of the set share the reachability information.

(4) Starting from the initial node u of the attack path graph, the breadth-first forward search algorithm is used to search for the node set V_i connected to the node u in the reachability matrix, If the intention is Intent $\in V_i$, then u is in reachable.

Using this model, it is only necessary to calculate the reachability of one of the interfaces in the filtering reachability domain between the subnets, which greatly reduces the calculation amount and time, and greatly improves the efficiency of the algorithm.

5. PROBABILITY CALCULATION BASED ON ATTACK INTENTION OF ATTACK PATH GRAPH

Through comparative study, this paper finds that Difficulty of vulnerability utilization, Gain of success and Stealths of attack can be used to quantify the probability of vulnerability utilization.

The attacker's attack level is divided into three levels: high, medium and low. The corresponding coefficients are [0.2, 0.4, 0.4], [0.5, 0.3, 0.2], [0.8, 0.0, 0.2]. The attacker's attack level can be dynamically adjusted by the security administrator based on the actual performance of the attacker. Then the vulnerability utilization probability V is:

$$V(v_i)=[\lambda_1, \lambda_2, \lambda_3] \cdot [D,S,G]^T \quad (1)$$

The assignment of the three factors of vulnerability quantification is shown in Table 3 below.

Table 3 Vulnerability attribute quantization table

Vulnerability attribute	Description	Assignment
Difficulty	Easy (there are ready-to-use attack tools and methods)	0.8
	General (no ready-to-use attack tools but attack methods)	0.5
	Difficulties (no off-the-shelf attack tools and no attack methods)	0.1
Stealths	Hard to be discovered when used	0.05
	May be discovered when used	0.5
	Must be discovered when being used	1.0
Gain	Larger (get important server root or user privileges)	1.0
	Medium (get the user permission of the common host)	0.8
	Smaller (general information disclosure)	0.5
	Very small (get host survivability information, etc.)	0.2

In order to achieve the attack intention, the attacker must ensure that all the vulnerabilities on a path from the current node u to the attack intention node v can be successfully utilized. Based on the probability V of a single vulnerability utilization, the total probability C that the entire attack path can complete can be calculated. Assume a complete attack path: $p_k = (u, vuln_1, \cdots, vuln_n, v)$ then:

$$\mathrm{C}(\mathrm{p_k}) = \mathrm{V}(\mathrm{v_1}) \cdot \mathrm{V}(\mathrm{v_2}) \cdots \mathrm{V}(\mathrm{v_n}) = \prod_{1}^{\mathrm{n}} \mathrm{V}(\mathrm{v_i}) \qquad (2)$$

The path from the initial location to the attack intent node is not necessarily unique, assuming there are m attack paths from node u to attack intent node v: $\{p_k, \mathrm{k} = 1,2,\cdots,\mathrm{m}\}$,In these m paths, as long as any one of them can be successfully completed, the attacker can achieve its final attack intention. Therefore, the attack path is a "noisy-OR" relationship, and the attack intent realization probability (A) can be calculated by:

$$\mathrm{A}(Intent_i) = 1 - \prod_{k=1}^{m} (1 - \mathrm{C}(p_k) \quad) \qquad (3)$$

6. EXPERIMENT

In order to verify the model and algorithm of this paper, the experimental environment shown in the following figure was built. The network consists of 3 domains DMZ, D1 and D2. In the domain DMZ, all three servers are directly connected to the Internet. All hosts in the D1 domain can access the servers in the DMZ domain. H4 in the D1 domain can access the SQL server in the D2 domain. Other cross-domain connections are limited by the firewall.

Use the vulnerability scanning tool to scan the experimental network, and initially determine the level of the attacker as the primary level, which can be adjusted according to the actual situation. The probability V that the vulnerability is utilized according to Equation 5 is utilized. See Table 4 for the vulnerability of the host in the experimental environment:

Table 4 List of vulnerability information in the experiment

Host	OS	Vulnerability identification	CVE number	V
H_1	VPN-1 Server 4.1	v_1	CVE-2004-0040	0.84
H_2	Windows 2003 Server	v_2	CVE-2006-2379	0.74
H_3	RedHat Linux 9.0	v_3	CVE-2003-0252	0.74
H_4	Windows XP	v_4	CVE-2004-0575	0.80
H_5	Windows 2003 Server	v_5	CVE-2008-0702	0.56
H_6	Windows XP	v_6	CVE-2004-0893	0.84
	SQL	v_7	CVE-2003-0004	0.84
H_7	Windows 2000	v_8	CVE-2007-0038	0.60
H_8	Windows XP	v_9	CVE-2006-2370	0.74

According to the vulnerability information in the host in Table 4, combined with the algorithm steps in 4, generate a vulnerability attack path map in the network, and calculate the probability that each vulnerability is successfully utilized. As shown in Figure 5.

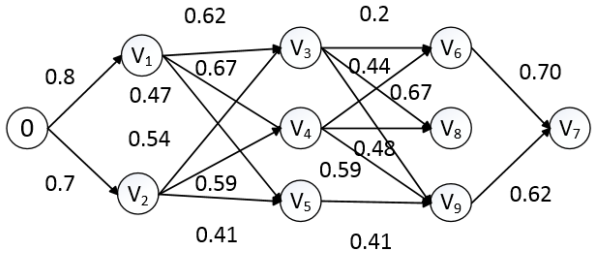

Figure 5 Vulnerability attack path graph

According to the host where the vulnerability is located and the connection relationship between the hosts, the attack path of the host level during the attacker attack is obtained, as shown in Figure 6.

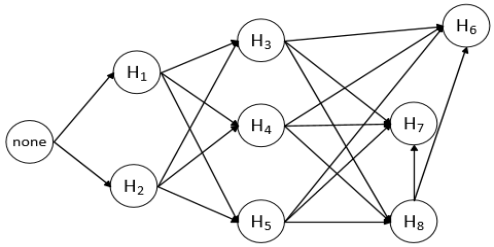

Figure 6 Host attack path diagram

From the experimental network topology, combined with the host-level attack path map, the management domain-level attack path map is： Internet $\longrightarrow$ DMZ $\longrightarrow$ D1 $\longrightarrow$ D2.

According to the above attack path map, the possible attack paths of important asset hosts H6 in the experimental environment and the probability of implementation of each path can be obtained as shown in Table 5 below:

Serial number	Vulnerability level attack path	Host-level attack path	Achieve probability
1	v_0->v_1->v_3->v_6	H_0->H_1->H_3->H_6	0.52
2	v_0 -> v_1 -> v_3 -> v_9 -> v_7	H_0 -> H_1 -> H_3 -> H_8 -> H_6	0.39
3	v_0 -> v_1 -> v_5 -> v_9 -> v_7	H_0 -> H_1 -> H_5 -> H_8 -> H_6	0.29
4	v_0 -> v_1 -> v_4 -> v_9 -> v_7	H_0 -> H_1 -> H_4 -> H_8 -> H_6	0.42
5	v_0->v_1->v_4->v_6	H_0->H_1->H_4->H_6	0.56
6	v_0->v_2->v_3->v_6	H_0->H_2->H_3->H_6	0.46
7	v_0 -> v_2 -> v_3 -> v_9 -> v_7	H_0 -> H_2 -> H_3 -> H_8 -> H_6	0.34
8	v_0 -> v_2 -> v_4 -> v_9 -> v_7	H_0 -> H_2 -> H_4 -> H_8 -> H_6	0.37
9	v_0->v_2->v_4->v_6	H_0->H_2->H_4->H_6	0.50
10	v_0->v_2->v_5->v_9->v_7	H_0 -> H_2 -> H_5 -> H_8 -> H_7	0.26

After the attack intention is to steal the SQL data information in the host H_6 as the attack intention Intent, it can be seen from the above table that the attack path with the highest probability of occurrence is the attack path No.5, that is, the attacker is most likely to take the attack path No.5. However, there is more than one path to achieve the intent of the attack. As long as one of them succeeds, the attack intention can be achieved. Therefore, from the calculation of equation (3),the probability of implementing the attack intention Intent is 0.89. It can be seen that in this network environment, the important asset is very vulnerable to attack. It can be seen from the above attack diagram that the communication between the host H_8 and the host H_4 and the host H_6 should be immediately cut off.

7. CONCLUSION

We propose a method based on the hypothesis of critical asset attack intentions. The important assets in the network system are taken as the assumed attack target, and corresponding protection strategies are formulated, which effectively protect the critical assets in the system from or less attacks. On the basis of the vulnerability causal model, a hierarchical attack path map is constructed. The attack graph model has good expansibility and can be well applied to the changes of network environment.It can also realize the analysis of network state finalization and quantification. In the attack path diagram, the probability of each assumed intention target is calculated by reasoning, and corresponding protection measures are formulated to provide data support for the formulation of network security protection strategy. Experimental results show that the proposed algorithm in this paper is feasible.

At present, we have carried out a series of explorations on the analysis of attack intentions, and have achieved some preliminary results. There are still some problems to be solved. For example, the occurrence of cyberattacks is time-sensitive, and how to analyze and identify the attacker's attack intentions in real time and dynamically will be further improved in future work.

8. ACKNOWLEDGMENTS

This research work has been supported by the National Natural Science Foundation of China (No.61300103), the Technology Innovation Platform Project of Fujian Province(No.2009J1007), the Science and Technology Project of Fujian Province,China(No.2014H0024,2015H6013), Corresponding Author: Prof. Minchen zhu, zmc@fzu.edu.cn.

9. REFERENCES

[1] Zhuang R, Bardas A G, Deloach S A, et al. A Theory of Cyber Attacks[C]// ACM Workshop. ACM, 2015:11-20..

[2]Lingyu Wang, A.S.S.J.. Measuring the Overall Security of Network Configurations Using Attack Graphs. Lecture Notes in Computer Science, 2007. 4602: p. 98-112.

[3]Bratman M. Intentions,Plans,and practical Reason [M].Harvard University Press,MA,1987.

[4]Ding H, He Q, Zeng L, et al. Motion intent recognition of individual fingers based on mechanomyogram[J]. Pattern Recognition Letters, 2017, 88(C):41-48.

[5]Russell D,Gangemi G. Computer Security Basics [M].Sebastopol,CA:O'Reilly &Associates,Inc.,1991:9-10.

[6] Bopche G S, Mehtre B M. Attack Graph Generation, Visualization and Analysis: Issues and Challenges[C]// International Symposium on Security in Computing and Communication. Springer, Berlin, Heidelberg, 2014:379-390.

[7]Kaynar K, Sivrikaya F. Distributed Attack Graph Generation[J]. IEEE Transactions on Dependable & Secure Computing, 2016, 13(5):519-532.

[8] Tiwari V K, Dwivedi R. Analysis of cyber attack vectors[C]// International Conference on Computing, Communication and Automation. IEEE, 2017..

[9] Kim S, Lee H, Seo H, et al. A Study on the Intention Analysis of Cyber Attack[J]. 대한전자공학회 학술대회, 2017.

[10] Kościelny J M, Syfert M, Wnuk P. The Idea of On-line Diagnostics as a Method of Cyberattack Recognition[J]. 2017.

[11] Oliver D,Cunningham R K.Fusing a Heterogeneous Alert Stream into Scenarios[A].Proc of the 2001 ACM Workshop on Data Mining for Security Applications[C].2001 1-13

[12] Ahmed A A, Mohammed M F, Ahmed A A, et al. SAIRF: A similarity approach for attack intention recognition using fuzzy min-max neural network[J]. Journal of Computational Science, 2017.

[13]Boosting performance in attack intention recognition by integrating multiple techniques[J]. China's computer science frontier,2011,(1):109-118. DOI:10.1007/s11704-010-0321-y.