An Atomic-Domains-Based Approach for Attack Graph Generation

Fangfang Chen, Chunlu Wang, Zhihong Tian, Shuyuan Jin, Tianle Zhang

Abstract—Attack graph is an integral part of modeling the overview of network security. System administrators use attack graphs to determine how vulnerable their systems are and to determine what security measures to deploy to defend their systems. Previous methods on AGG (attack graphs generation) are aiming at the whole network, which makes the process of AGG complex and non-scalable. In this paper, we propose a new approach which is simple and scalable to AGG by decomposing the whole network into atomic domains. Each atomic domain represents a host with a specific privilege. Then the process for AGG is achieved by communications among all the atomic domains. Our approach simplifies the process of design for the whole network, and can give the attack graphs including each attack path for each host, and when the network changes we just carry on the operations of corresponding atomic domains which makes the process of AGG scalable.

Keywords—atomic domain, vulnerability, attack graphs, generation, computer security

I. INTRODUCTION

As networks of hosts and security incidents continue to grow, it becomes increasing more important to automate the process of evaluating network security. Attack graphs are composed by all the attack paths which lead to intruders intentions. The attack path is formed by a chain of exploits, where each exploit is realized by taking advantage of known vulnerabilities in various of services and systems. The term vulnerabilities refers to exploitable errors in Configurations and server software implemented to provide network services. Essentially, attack graphs represent the security state of network, and can serve as an useful tool in several areas of network security, including intrusion detection, defense, and forensic analysis[8]. On the one hand, system administrators can use attack graphs to collect information about their system’s security state[9][18][23]. On the other hand, minimal-critical sets of vulnerabilities and key hosts or vulnerabilities can be computed by attack graphs[7]. Then measures can be made to strengthen the network security. This active defense can achieve better effect than the passive defense which is achieved by collecting succeed attacks.

Supported by the National High-Tech Research and Development Plan of China under Grant Nos. 2009AA01Z438, 2009AA01Z431; the National Natural Science Foundation of China under Grant No.60703021; Fangfang Chen is with the Department of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China e-mail: luyun_2008@163.com.

Chunlu Wang, 1969, Beijing University of Posts and Telecommunications, Associate Professor, Computer Architecture and Computer Security, wangcl@bupt.edu.cn

Zhihong Tian, Ph.D, Computer Network, Computer Science, tianzh@bupt.edu.cn

Shuyuan Jin, Ph.D, Computer Security, jinshuyuan@software.ict.ac.cn

Tianle Zhang, Ph.D, Computer Network, tianle Zhang@bupt.edu.cn

Most of methods for AGG are based on privilege promotion [1][6][7][8] and information used to describe pre- and post-conditions for exploitors attack components must be entered by hand. This is labor intensive and difficult [20]. Now, the Common Vulnerability Scoring System (CVSS) provides an open framework for communicating the characteristics and impacts of IT vulnerabilities [17]. There are many descriptions for vulnerabilities’ pre- and post-conditions. But for generating attack graph, there is an essential characteristic called relationship between pre- and post-conditions, then vulnerabilities post-conditions can be part of another vulnerabilities pre-conditions. Although some other vulnerability database also provide information relation to pre- and post-conditions, they do not contain the machine-readable details required to accurately produce many of the attacks graphs shown in past papers [20]. The lack of study on vulnerability relatedness has been the bottleneck in attack graphs generation, analysis and usability.

The generation of attack graphs has experienced two stages including manual generation and automatic generation[7][8]. In automatic generation, model checking owns the most widespread application and approval[1][6][22]. (Such as the usage of model checking tool called NuSMV[7][8][9][22]) Besides, logic programming is also demonstrated to be an effective method owning to attack graphs automatic generation[4]. But for practical application, there are several shortcomings as follows in methods above: 1. Non-automated: For model checking, there is no transformation tool for the networks automatic modeling[7][8][9][22]. For modern complex and large-scale network, this manual network modeling is labor-intensive, difficult, and error-prone. 2. Incomplete: In model checking, the attack graphs are aiming at some host[7][8]. In fact, the administrator is not sure about which host is needed to be protected, so it is essential to generate the attack graphs for the whole network. 3. Complex: All the present methods for attack graphs generation including model checking and logic programming are aiming at the whole network during operation. It is quite complex to deal with the information of the whole network including all the hosts, network topology information and vulnerabilities. 4. Non-scalable: Using the present methods, the scalable generation is not achieved. When information of network are changed (such as some hosts are added or removed) then processes for generation are again. 5. Impractical: Almost all the network topology discovery tools can just achieve the getting of topology information concerning some specific host[19][20]. This makes the process of the whole network information first, generation second.
The method of this paper provides away by which

the attack graph can be generated without the whole network topology information. We just need the realization of getting topology information concerning some host.

This paper describes a new approach to attack graphs generation which is concern about the information design of each host, but not the whole network. This method is essentially a re-distribution of the information of network. Our method has the advantages as follows:

1. Simplify the process of attack graphs generation: Comparison with the original methods which is dealing with the whole information of network, we are aiming at the information of each host. Besides, the design for each host is in the same template.

2. Gain the attack graphs for all the hosts of the target network: When analyzing the security of an enterprise network, it is important to consider multi-host attacks. At present, there are hand generated, programming with some programming languages and so on. But, the methods above have poor scalability and are limited in practical applications. Our approach has the ability to multiple goals attack graph generation and has good scaling.

3. Achieve the generation of attack graphs directly without the concern with the original methods which is dealing with the whole network. Our method can achieve the goal without the information of attack paths. And if necessary, attack paths can be obtained by analyzing attack graphs.

II. RELATED WORK

Philips and Swiler propose the concept of attack graph and describe a tool for attack graphs generation. After that, a variety of approaches to generating various forms of attack graphs appear including custom-design graph-based and model checking. In model checking method, the attack paths are given as counter-examples which are related with the safety policy given by network administrators. NuSMV is the improvement of SMV and all the counter-examples that violates the safety policy can be achieved. In the prior network models, the counter-examples are attack paths corresponding in the safety policy that the intruder can not earn some privilege on some host. By integrating all the attack paths, the attack graph is generated. Different from prior work, our approach aims at generating the attack graph for the whole network. In other words, we integrate all the attack paths for all the hosts in the target network. Past attack graphs are subsets of our attack graph and attack graphs for each host can be earned by reverse depth-first search from target host. There are various of attack graphs with different kinds of nodes or edges. In paper, the nodes in their attack graphs represent the state of the network, and the edges represent attacker actions that change the state. While in [10], their nodes of attack graphs represent an host and the edges represent the vulnerability used to attack next host. Taking into account the complexity of attack graphs, the former attack graphs have too many nodes, while the latter one has too many edges on each nodes. In paper, the authors give one design which is the compromising position of the two illustrations above. In the initial stage, there is no limiting factors and the attack graphs is complete that means all the attack paths are included. Then the significant exponential explosion problem happens. Ammann, et al. pointed out that for most computer attacks, one can assume the monotonicity property, where an attacker does not decrease his ability by launching attacks, and hence does need to relinquish privileges he already gained. Under this assumption, an attackers privileges always increase during the analysis. Since there are only a polynomial number of privileges an attacker can gain, the analysis algorithm will terminate in polynomial time. Our approach achieves this by designing of atomic domains. In paper, the authors use an efficient semantic evaluation program in the MulVAL reasoning engine to generate the logical attack paths. They got better running time than Ammanns. After that, Rattikorn Hewett and Phongphun Kijsanayothin got even better running time than paper by host-centric model checking. This paper achieve even better improvement than by eliminating unnecessary internal attacks in some host.

III. ATTACK MODEL

A. Definitions

An attack graph model is an attack model where all informations concerning network including network topology, hosts, vulnerabilities and so on are organized to be a finite state machine whose state transitions are based on intruders attack actions. Former approaches for attack graphs generation are aimed at the whole network. In this paper, our approach divided the network modeling into design of each host. The generation of attack graphs can be achieved by communications between atomic domains. Here are some definitions for our system:

Definition 1. $V ul = \langle ID, pre, post, h, s >$ is a five-tupled definition for a vulnerability. Each vulnerability has five properties as follows:

1. $V ul . ID$, the CVE standard vulnerability name
2. $V ul . pre$, intruder pre-condition
3. $V ul . post$, intruder post-condition
4. $V ul . h$, the host owning the current vulnerability
5. $V ul . s$, a boolean and when valued 1(true), it means the intruder has been succeeded in capitalizing on the current vulnerability.

Definition 2. An $AD h = \langle h, n, VN, VA, t >$ is an atomic domain for host $h$ with privilege. The following is a list of the components for each $AD h$:

1. $h$, a host with the name $h$
2. $n$, the privilege for $h$
3. $s$, the selection for $n$. There are three values for $n \in (0, 1, 2)$ representing (none, user, root).
3. VN(Vulnerabilities Needed), a set of vulnerabilities owning to the host or hosts which are connected to the host directly and making it possible to achieve privilege on the host in an exploit.

4. VA(Vulnerabilities Available), a set of vulnerabilities owning to the host or hosts which are connected to the host directly and being available to use when the intruder achieves privilege on the host.

5. t, a boolean variable. When valued 1(true), it means the intruder has achieved the privilege n on the host. Then attacks concerning the set of vulnerabilities VA can be started.

**Definition 3.** An attack is achieved by communication between atomic domains and the communication is achieved by two actions as follows:
1. $AD_n^h \cdot a(vul)$: sending action performed by $AD_n^h$ whose vulnerabilities set VA owns vul.
2. $AD_n^h \cdot a(vul)$: receiving action performed by $AD_n^h$ whose vulnerabilities set VN owns vul.

The two above are boolean and mean the success of exploit concerning vul when both valued 1(true).

**Definition 4.** Success label: we associate a boolean function for each vulnerability, abstractly representing whether or not the current vulnerability is used successfully. For the vulnerability Vul, we define the function as $a^2(Vul)$.

**Definition 5.** In each attack path such as $(s_1, s_2, \ldots, s_n)$ where $s_i$ indicates an $AD_n^h$ intruded during attacking, $(s_i, s_{i+1})(1 \leq i \leq n)$ is called AA (atomic attack). AA has two elements as follows:
1. $s_1 \cdot st$ (single source), the attack sponsor in an AA. If $(AA = (s_1, s_{i+1}))$ Then $AA.s = s_i$.
2. $s_1 \cdot st$ (single target), the attack goal in an AA. If $(AA = (s_1, s_{i+1}))$ Then $AA.st = s_{i+1}$.

**Definition 6.** $NAD$(Neighbor Atomic Domains) of a host $h$ is a set of $AD_n^h$, in which the $AD_n^h$ owning to the host which is reachable to the current host $h$ directly.

**Definition 7.** An attack model is a finite automaton $M = (\langle AD_n^h, s, a^2s \rangle)$, where $AD_n^h$ is a set of all the $AD_n^h$ owning to the network and $a^2s$ is a set of success labels for all vulnerabilities owning to the network. The value changes of the set $a^2s$ mean the state transformation of the finite automaton.

**B. Modeling**

The process for network modeling is as follows:
1. To store the hosts newly discovered each time, we create a queue called Q.
2. Using network topology discovery tool, we get the set $C$ of hosts which are reachable directly to the attacker and store the hosts of $C$ into the queue Q. Then the vulnerabilities information of hosts in $C$ can be got by the vulnerability scanning tool Nessus. Finally, the initialization of $AD_n^h$ (atomic domain) concerning attacker can be achieved. Take a host from Q, and go with the following steps:

1) Using the network topology discovery tool, we get the set $C$ of hosts which are reachable directly to the current host.
2) Queue the hosts newly discovered in (1) into Q.
3) Get the vulnerabilities information of each host from $C$.
4) Initialize the $AD_n^h$ (atomic domain) concerning the current host.
5) Check whether the queue Q is empty. If Q is empty, the modeling for network is achieved. Otherwise, go back to 3.

1) Parameter Settings: In accordance with the definition in 2.1, for the target network with n hosts, there should be $3(n + 1)$ $AD_n^h$s (including three $AD_n^h$s concerning the intruder). But for most computer attacks, an attacker follows the monotonicity property, where he dose not decrease his ability by launching attacks and will focus on privilege promotion [4]. This property can be valuable to the complexity reduction of attack graphs. During setting of $AD_n^h$s, we achieve monotonicity property by rules as following:

1. In the initial state of the network, as the intruder owns the privilege of administer on its own, it is not needed to notice the other two lower privileges. Then only one atomic domain is necessary for the host a owned by the attacker: $AD_2^n$.
2. In the initial state of the network, as the intruder owns at lest the privilege of none on each host of target network, there is no necessary to care about the privileges of none for each host. The intruder can achieve the privilege of none on each host without attacks. Then only the $AD_n^h$s concerning the privileges of user and administer on each host are necessary. Above all, $(2n + 1)$ but not $3(n + 1)$ $AD_n^h$s are necessary in our design. We name the hosts of the target network as $(1, 2, \ldots, n)$. Then the set $AD_n^h$s owning the attack model M represents a set of $(2n + 1)$ agents $AD_n^h = (AD_1^n, AD_2^n, AD_3^n, \ldots, AD_n^n)$. In which $(AD_1^n, AD_2^n)(1 \leq i \leq n)$ are the $AD_n^h$s concerning the host $i$.

2) Initialization of $AD_n^h$: For atomic domain $AD_n^h$, the initialization of the properties $AD_n^h.VN$, $AD_n^h.VA$, $AD_n^h.t$ concerning $AD_n^h$ is as follows:

1. $AD_n^h.VN$: The attack towards host $h$ can be achieved just by exploits concerning its own vulnerabilities. Then elements of $AD_n^h.VN$ are from host $h$ and another condition should be: $Vul.post = n$.
2. $AD_n^h.VA$: The vulnerabilities available for $AD_n^h$ can either come from host $h$ or hosts which are reachable to $h$ directly. Then the setting of $AD_n^h.VA$ is as follows:

1) For the vulnerabilities owning host $h$, the one which meets the rule: $Vul.pre = n$ and is not included in $AD_n^h.VN$ can be included in $AD_n^h.VA$. In our design, we prevent the redundant attacks by making sure there is no common elements in $AD_n^h.VN$ and $AD_n^h.VA$.
2) For the vulnerabilities owning hosts connected physically with host $h$, the one which meets the rule: $Vul.pre \leq n$ can be included in $AD_n^h.VA$. When gets higher privilege on a...
host, the intruder can achieve all the exploits concerning the vulnerabilities with lower pre-conditions.

3. $AD_n^h.t$: When existing one vulnerability $Vu_l$ in the set $AD_n^h.VN$ which meets the rule: $Vu_l.s = 1$, then $AD_n^h.t$ can be valued 1.

Among all the pre-conditions needed for attack graph generation, network topology is an important part, while in prior work this part is handled as a single and complete part [1][6][7][8][9][10][16][22]. Then before generation attack graphs, the whole information about network topology is needed. As it is hard for the current network topology discovery tools to get the network topology, current methods for attack graphs generation are impractical for unknown network. In our approach, only the topology information concerning one host is needed and it is easy for topology discovery tools to achieve this. In our approach, we include the network topology information into the setting of $AD_n^h.VN$ and $AD_n^h.VA$.

C. Algorithm

After setting and initialization of $AD_n^h.s$ (atomic domains), the preparations for attack graphs generation have been achieved, then the process of generation can be achieved by communications between $AD_n^h.s$.

Actually, our method gives the process of breadth-first penetration attacks to network. In our design, we achieve the lamination attack graph generation by the data structure queue. To the attack model $M = (AD_n^h.s, a^2.s)$, in which the set $AD_n^h.s$ represents a set of $2^n + 1$ agents $AD_n^h.s = (AD_n^h, AD_n^h, AD_n^h, AD_n^h, AD_n^h, \cdots, AD_n^h, AD_n^h)$, we can achieve the generation of attack graphs as follows:

1. To achieve the breadth-first traversal of the elements in the set $AD_n^h.s$, we create a queue called $Q$ to store the $AD_n^h.s$ need to be activated.
2. Started from $AD_n^h$, the attacks concerning the vulnerabilities owing to $AD_n^h.VA$ can be started. Then the $AD_n^h.s$ being attacked are queued into $Q$.
3. Take one atomic domain from $Q$ (named $AD_n^h$), go with the following steps:
   1) Active $AD_n^h$.
   2) Start the attacks concerning vulnerabilities owning to $AD_n^h.VA$.
   3) Check the target $AD_n^h.s$ in (2), and queue the $AD_n^h.s$ which have not been activated into $Q$.
   4. Check whether the queue $Q$ is empty. If $Q$ is empty, the generation of attack graph is achieved. Otherwise, go back to 3.

D. Consideration of Implementation

SPIN is a generic verification system that supports the design and verification of asynchronous process system. Spin verification is focused on proving the correctness of process interactions, and they can simulate the communications between processes. In our design, each atomic domain is corresponded to a process. Then the communications between ADs can be simulated and achieved.

The modeling language for SPIN is Promela (Protocol/Meta Language). Fig. 1 shows the model for Promela.

IV. EXPERIMENT

Table 2 shows the information for communication stored in channels: ($sAD, tAD, Vu_l$).

Once there exits one vulnerability in $AD_n^h.VN$ which is exploited by intruder, $AD_n^h$ is activated by valuing the property $t$. Then the attacks concerning vulnerabilities in the set $AD_n^h.VA$ can be started by $AD_n^h$.
TABLE III
SUMMARY OF EXPLOITS

<table>
<thead>
<tr>
<th>Vulnerability/Trust Exploit</th>
<th>Victim host</th>
<th>Pre-con. On Attack host</th>
<th>Post-con. On Victim host</th>
<th>Exploit Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>tns Oracle TNS listen buff.Ovf</td>
<td>D</td>
<td>Access ≥ 1</td>
<td>Access = 2</td>
<td>Remote</td>
</tr>
<tr>
<td>t1 Trust Remote login(Any to W)</td>
<td>W</td>
<td>Access ≥ 1</td>
<td>Access = 1</td>
<td>Remote</td>
</tr>
<tr>
<td>t2 Trust Remote login(W to D)</td>
<td>D</td>
<td>Access = 1</td>
<td>Access = 1</td>
<td>Remote</td>
</tr>
</tbody>
</table>

A. Modeling for Example Network

According to the design, we can achieve the modeling for network in Fig. 2 as follows:
1. Setting of $AD_h^n$ owning to attack model $M$: $AD_h^n = (AD_2^n, AD_1^n, AD_2^n, AD_1^n)$
2. initialization of each atomic domain :Table IV
3. Encoding : this part is achieved by the language Promela.

B. Attack Graph Generation and Analysis

By SPIN, we can get the attack graph as shown in Fig.3, in which the processes (A2, W1, W2, D1, D2) are corresponded to $AD^n = (AD_2^n, AD_1^n, AD_2^n, AD_1^n)$. Different from prior study, our method achieved the generation of attack graph without knowing each attack path. We achieve the generation of attack graph first, then attack paths can be get by depth-first search to attack graph. As the goal of the study is the attack graph [1][6][7][8][16][22], it is not necessary to get attack paths. Our method generate the complete attack graph which includes all the attack path concerning each host. This is also different from previous works which are aimed at a certain known host [1][6][7][8][16][22].

V. COMPARISON

A. Applications

Consider the network attack graph shown in Fig. 4, which is called host-centric attack graph generated in paper [1]. Host-centric model checking achieved lower complexity than Network-centric and access graph [1]. But in Fig. 4, there appears exploits of privilege reduction in one host's internal attacks (such as attack from (W, 2) to (W, 1)), which does not meet the rule of monotonicity property. In our methods, we avoid this kind of situation by definition and design of atomic domains.

Fig. 5 gives a new network which is achieved by adding two hosts on the network in Fig.2. The vulnerabilities for the...
TABLE V
SUMMARY OF EXPLOITS

<table>
<thead>
<tr>
<th>Vulnerability/Trust Exploit</th>
<th>Victim host</th>
<th>Pre-con. On</th>
<th>Post-con. On</th>
<th>Exploit host</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>licq</td>
<td>W</td>
<td>Access ≥ 1</td>
<td>Access = 2</td>
<td>Remote</td>
<td></td>
</tr>
<tr>
<td>Exploit a problem in the URL parsing function of the LICQ software for Unix-flavor system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>se</td>
<td>D</td>
<td>Access ≥ 1</td>
<td>Access = 1</td>
<td>Remote</td>
<td></td>
</tr>
<tr>
<td>The scripting action lets the intruder gain user privileges on Windows machines.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lb</td>
<td>W</td>
<td>Access ≥ 1</td>
<td>Access = 2</td>
<td>Remote</td>
<td></td>
</tr>
<tr>
<td>Local buffer overflow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wu</td>
<td>D</td>
<td>Access ≥ 1</td>
<td>Access = 2</td>
<td>Remote</td>
<td></td>
</tr>
<tr>
<td>WuFtpd socketprintf buff.Ovf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE VI
ELEMENTS OF EACH ATOMIC DOMAIN

<table>
<thead>
<tr>
<th>VN</th>
<th>VA</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD_{2}^n</td>
<td>ap,t1</td>
<td>1</td>
</tr>
<tr>
<td>AD_{1}^n</td>
<td>ap,tns,t2</td>
<td>0</td>
</tr>
<tr>
<td>AD_{2}^*</td>
<td>ap,t1,licq,se,lb,wu</td>
<td>0</td>
</tr>
<tr>
<td>AD_{2}^*</td>
<td>ap,licq,se,lb,wu</td>
<td>0</td>
</tr>
<tr>
<td>AD_{1}^*</td>
<td>licq,tns,t2</td>
<td>0</td>
</tr>
<tr>
<td>AD_{1}^*</td>
<td>se,tns,t2</td>
<td>0</td>
</tr>
<tr>
<td>AD_{2}^*</td>
<td>lb,tns,t2</td>
<td>0</td>
</tr>
</tbody>
</table>

new network are shown in Table 5.

According to the rules of ADs setting, the design for the atomic domains which are concerning about the two added hosts is shown in Table 6. And the attack graph for the new network is shown in Fig.6.

B. Discussion

Prior studies on attack graph generation focus on the whole complex network. Once the network topology changes, the target network must be re-modeled. Our method provides a way, in which we only focus on the tiny network concerning some host. This makes the modeling simple, scalable and reusable. This means that, we just focus on the related atomic domains when the topology or vulnerabilities change. Then the non-correlated part can be reused in the second modeling.

VI. SUMMARY AND FURTHER WORK

In view of present research status to attack graph generation, we propose a new approach which is simple and scalable by made of AD_{h}^{s}. On the one hand, we just pay attention to each tiny network concerning some host. On the other hand, when changes happen in network we just focus on the related AD_{h}^{s}. The most important is that we achieve the attack graphs which include all the attack graphs to every host owning to the target network and attack graphs for each host can be earned by reverse depth-first search from target host in our attack graph.

There is some room for improvement in our design because of the following shortcomings:
1. Information redundancy: there exits information overlapping between AD_{h}^{s}, because one vulnerability simultaneously belongs to one AD_{h}^{s}.VN property and another AD_{h}^{s}.VA property.
2. Execution randomness of processes in SPIN: the randomness execution of processes in SPIN makes trouble during encoding and we have to adjust the execution order of all the AD_{h}^{s}.

Our future work includes information adjustment of AD_{h}^{s} and try to different tools to achieve the AD_{h}^{s} orderly implementation.
REFERENCES


