Scenario Recognition based on Collaborative Attack Modeling in Intrusion Detection

Xuejiao Liu, Debao Xiao, Ting Gu and Hui Xu *Student Member IEEE*

Abstract— Recently, intrusion detection products have become widely available, and are beginning to gain acceptance as a worthwhile investment for network security. However, traditional intrusion detection systems (IDSs) only focus on low-level attacks and raise alerts independently, though there may be logical connections between them. At the same time, the amount of alerts becomes unmanageable including actual alerts mixed with false alerts. To address that problem, several approaches for alert correlation and attack modeling have been proposed these years. In this paper, we suggest collaborative attack modeling of general attack pattern for constructing multistep attack scenario, based on attack classification. The purpose is then to enable attack-attribute aggregation that make low-level alerts to high-level aggregated ones from heterogeneous IDS systems. In order to better construct complete scenario, causal correlation based on time series and statistical analysis is introduced to facilitate scenario recognition. Through the experimental results with DARPA Data Sets 2000 from Lincoln laboratory, it demonstrates the potential of the proposed approach as well as the effectiveness of our techniques.

Index Terms—**alert correlation, attack modeling, collaborative**

I. Introduction

In recent years, organizational dependence on networked information technology and its underlying infrastructure has grown explosively. In conjunction with this growth, the frequency and severity of network-based attacks have also dramatically increased [1]. To guard against these malicious attacks for network security, Intrusion Detection Systems (IDSs) are important and indispensable device being part of security mechanisms. They monitor protected network and attempt to identify evidence of malicious activity [2]. When an attack is detected, an alert is produced.

To be effective enough for network security, the ideal Intrusion Detection Systems (IDSs) must satisfy TAC principle. They are (a) ***timely***: when identifying spurious and malicious attack in a network, intrusion detection should be rapid enough to report and defense in time. (b) ***accurate***: the alerts triggered by intrusion detection should be successfully true events, and (c) ***complete***: intrusion detection should infer as complete a set of anomalies that can explain all the detected events as possible.

Unfortunately, traditional IDS systems only focus on low-level attacks and raise alerts independently, though there may be logical connections between them. Moreover, there are many other problems arising such as alert flooding and false alerts. The problem is even more pressing as how to identify the camouflaged intrusion more accurately from a huge amount of alerts. Thus, it is necessary to construct high-level attack scenarios from a large collection of low-level intrusion alerts [14].

To overcome limitations of traditional IDSs and enhance the efficiency of intrusion detection, researchers and practitioners attach great importance to **collaboration** in intrusion detection. One major challenge in collaborative intrusion detection is attack modeling of multiple alerts from various IDSs. The better attack knowledge base is established, the more likely attack scenario is discovered.

In this paper we have developed collaborative attack modeling of general attack pattern for constructing multistep attack scenario, based on attack classification. The purpose is to enable attack-attribute aggregation that make low-level alerts to high-level aggregated alerts from heterogeneous IDS systems. In order to better construct complete scenario, causal correlation based on time series and statistical analysis is introduced to benefit scenario recognition.

The remainder of the paper is organized as follows. In Section II, related work is discussed. Section III presents collaborative attack modeling in detail, including attack classification and general attack pattern. In Section IV, attack-attribute aggregation and casual correlation based on time series and statistical analysis is thoroughly presented. Then to verify the correlation suitable for describing attack scenarios, we implement scenario recognition with typical datasets in Section V. At last, we outline our conclusions and point out some future work in Section VI.

II. Related Work

Realizing the limitations of traditional IDS systems, researchers began to explore new ways to improve IDS performance. These years, some exciting and important advances have been made on scenario recognition.

Alert correlation: As intrusion alerts only reflect elementary steps in an attack, alert correlation methods aim at reconstructing the attack scenario by linking alerts that satisfy certain relationships together. Exemplary alert correlation work includes alert aggregation and correlation algorithm

This was part of the project Unified Platform of Security Management (200710421130) supported by Science and Technology Bureau of Wuhan Municipality.

Xuejiao Liu, is with the Institute of Computer Network and Communication, HuaZhong Normal University, Wuhan, P.R.China, 430079. (e-mail: liuxuejiao@gmail.com).

Debao Xiao, is with the Institute of Computer Network and Communication, HuaZhong Normal University, Wuhan, P.R.China, 430079. (e-mail: dbxiao@mail.ccnu.edu.cn).

Ting Gu, is with the Institute of Computer Network and Communication, HuaZhong Normal University, Wuhan, P.R.China, 430079. (e-mail: tng_gu@yahoo.com).

Hui Xu is with the Institute of Computer Network and Communication, Huazhong Normal University, Wuhan, 430079, P. R. China (e-mail: xuhui_1004@hotmail.com).

[17], attribute similarity based approach [21] [22], clustering and merging function [16], a general correlation model [18], and pre/post condition based [12] [13] [14]. We extend a simple and flexible causal correlation based on time series and statistical analysis derived from [20] in pre/post matching. Attack-attribute aggregation can group alerts from multiple IDS systems with appropriate granularity in scenario constructing.

Attack Modeling: For detecting multi-step attack scenarios, attack modeling becomes quite important to facilitate the analysis of intrusion alerts. Several attack modeling approach has been proposed in M2D2 [5], LAMBDA [15], CAML [11], TIAA [12], a logical formula *capability* [6], a fault tree-like method [8] [4] and a graph-based technique with a Web-based collaboration tool [19]. We provide a finer abstraction of IDS alerts considering different attack types, which is effective and efficient in alert correlation.

III. Collaborative Attack Modeling

Attack scenario is a sequence of steps taken by the intruder, who typically culminating in a particular goal——administrative access on a particular host, Denial of Service, and etc. As the attacker's goals and methods play the central role in nearly all security considerations, attack models are closest to the problem and therefore are useful for the discovery of attack scenarios.

An elementary attack corresponds to a non-decomposable step of a given scenario. Attack scenario modeling is related to attack graphs used by attackers. However, the purpose of the attack models is not only to provide details on how each attack is to be carried out, but also to aggregate the various alerts on how the attacks are detected and reported.

To identify the actions and steps taken by the attacker in a proactive manner effectively, it is quite important to structuring multiple alerts and providing a formal model to reason and construct possible scenario of attacks referring to IDS alerts.

A. Attack Classification

There has been some attempt to attack classification. A classification of attackers could be made based on their capabilities, resources, or motivation. Our solution to the problem is collaborative attack modeling, which lies in the creation of a vendor-independent attack representation - one that can be converted to IDS-specific representations or models automatically.

In order to make the results even more comprehensible, we categorize the target into five different classes {Probe，Buffer Overflow, Compromises, DOS, Worm/Trojan horse}.

The detailed description is specified as follows.

Probe denotes pinging, probing and scanning, such as host or port scanning [10]. An attacker may probe to collect the list of valid IP addresses within a network, and the basic information about the system such as OS, the services it runs, the port it opens and so on.

Buffer Overflow denotes storing more data in a buffer than it was intended to hold. Since buffers are created to contain a finite amount of data, the extra information which has to go somewhere, may overflow into adjacent buffers, corrupting or overwriting the valid data held in them, or even trigger specific action to the computer.

Compromises denote using known vulnerabilities such as buffer overflows and weak security to gain privileged access to hosts. There are two ways of comprises. One is ***R2L*** (Remote to Login), which refers to unauthorized access from a remote machine, such as guessing a password. The other is ***U2R*** (User to Root), which means unauthorized access to local super user privileges, such as various buffer overflow attacks.

DoS (**Denial-of-Service**) attacks (e.g., smurf, land) usually attempt to shut down a network, computer, process and otherwise deny the use of resources or services to the authorized users.

Worm/Trojan horse is well-know to us, which can aggressively replicate on other hosts. The difference between worm and Trojan horse is that worms are self-replicating, while Trojan horses are downloaded by users.

B. Collaborative Attack Patterns

Developing attack models for multistep attack scenarios could be quite time-consuming [11]. Moreover, corresponding to a specific attack (e.g., DDoS attack), different IDS systems trigger multiple alerts, although many are duplicate ones, some are false positives, and there are even logical correlation among them. Thus, it is important to identify methods for building new attack models based on collaborative intrusion detection.

Attack patterns facilitate attack model reuse. These attack patterns correspond to high-level reusable modules that characterize common attack techniques from the detection point of view. Generally, there are three necessary phases in a successful attack.

Phase 1 Planning phase Before making an attack, an attacker often make use of the system in its intended manner via different forms. The motivation behind the attack determines the forms. Often an attacker may have goals such as Denial of Service, escalation of legitimate privileges, unauthorized access or data manipulation, and etc. After the initial preparation is complete, the attacker decides on the scope of the attack.

Phase 2 Reconnaissance phase The goal of the attacker in this phase is to narrow down the field of thousands of possible exploits to a small number of vulnerabilities that are specific to the targeted network. An attacker can get information through legitimated public data available in forums, public databases, public monitoring tools, or through vulnerability scanning methods such as ping, TCP connect, and OS version scanning and etc.

Phase 3 Attack phase After the previous preparation is completed, various attacks are ready to launch. Several attack ways are described as follows.

Denial of service: Any attack that disrupts the function of a system so that legitimate users can no longer access it.

Remote exploits: Attacks designed to take advantage of improperly coded software to compromise and take control of a vulnerable host.

Trojans and backdoor program sacks: Attacks to gain privileged unauthorized access to a host by installing a backdoor program or a Trojan and then bypassing normal security controls.

Misuse of legitimate access: Attacks often attempt to gain unauthorized use of legitimate accounts by getting a hold of authentication information.

IV. Scenario Recognition

Scenarios may be instantiated and recognized through alert aggregation and correlation mechanism in the use of knowledge-based interpretation. This allows the user to focus on the strongest scenarios, especially the final attack step.

A. Attack-attribute based Alert Aggregation

A common limitation of existing intrusion detection systems is that they normally issue too many alerts. The sheer number of IDS alerts can be overwhelming because current IDSs often trigger thousands of alarms per day. A lot of those alerts are duplicated in the sense that they are identical except that they are issued at a slightly different time, or issued by detectors installed at different locations. It's desirable to have as few **EFFECTIVE** alerts as possible when reporting the same ongoing attack.

In practice, one important variant that affects the results of aggregation is the evaluation of attack class. It is common to observe that multiple different attack classes are defined for attacks exploiting the same vulnerability or having similar result. To eliminate duplicate alerts and effectively group similar ones, we aggregate various alerts from multiple IDS systems based on attack-attribute.

Recently, researchers and practitioners attach great importance to **collaboration** among different IDSs (such as network-based IDS, host-based IDS as well as application-based IDS and etc). However, every IDS system has its own signature base with different granularity to detect intrusion and trigger alerts. For example, referring to Rsh, there is only one alert *Rsh* in RealSecure, while seven detailed alerts with different aspects are defined in Snort, such as *RSERVICES rsh bin*、*WEB-CGI rsh access*、 *EXPLOIT CVS rsh annotate revision overflow attempt* and etc.

In order to collaborate among different IDSs, a unified representation of alerts and their relationship are of great need. As trivial differences of various alerts do no use to scenario recognition, we consider the class of attack on a higher abstraction level illustrated in Table 1. For example, a “SNMP trap TCP” alert and a “SNMP trap UDP” alert generated by Snort can be merged into one hyper alert with the same abstracted class “SNMP trap”.

Table 1 Example of alert abstraction with Snort and RealSecure

Attack Type	Aggregated	Alert	Source
Probe	Sadmind Ping	RPC portmap sadmind request UDP	Snort
		Sadmind_ping	Realsecure
Buffer Overflow	Sadmind Overflow	RPC sadmind UDP NETMGT_PROC_SERVICE CLIENT_DOMAIN overflow attempt	Snort
		Sadmind_Amslverify_overflow	Realsecure
Compromises	Rsh Root Acess	RSERVICES rsh root	Snort
		Rsh	Realsecure
DoS	DoS	DOS Land attack/DOS ath	Snort
		Stream_DoS	Realsecure
Worm Trojan horse	Mstream	DDOS mstream agent pong to handler	Snort
		Mstream_Zombie	Realsecure

To support attack model extension (e.g., ability to incorporate new attack knowledge in attack models) and module composition, every alert is mapped into the specified type of the knowledge base in the aggregation process. Once a new alert referred to predefined attack type is defined into the signature base, only the mapping filed needed to add one up in the knowledge base. While if a novel attack type is emerged, the knowledge base is specified to update.

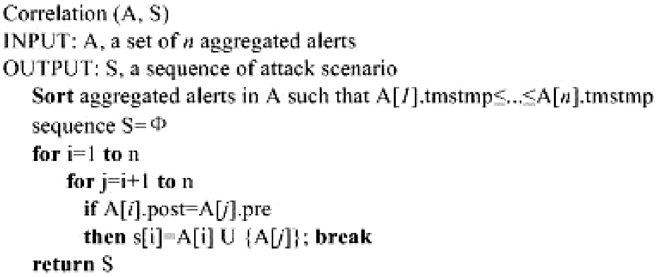
B. Causal Correlation based on Time Series and Statistical Analysis

By combining multiple IDS alerts into high-level alerts, the alert correlation can rapidly reduce the number of false alerts and suppress alert flooding.

As for alert correlation, the relationships between events can be classified as being ***causal*** (e.g., one attack enables another one to occur), ***temporal*** (e.g., one attack happens before another one) or ***spatial*** (e.g., one attack relates to another one in the network topology). As the main relationship for multistep scenario that can be considered, a causality relation or a cause-effect relation is to identify logical attacks in an attack scenario. And since events happen at particular time instants, richer correlations between events based on the specific instants of their occurrence can be established by defining temporal relationships between them. Additionally, spatial-based correlation correlates alerts from multiple observation spaces or sensors at the same time to detect attack scenarios.

To better express these correlation relationships, we conduct causal correlation based on time series and statistical analysis to specify inference step for scenario recognition. In order to integrate aggregated alerts for different types of attack, a unified pattern representation form at is of great need to discover attack sequences and recognize attack scenarios. Thus we design the correlation knowledge base as **Attack type_aggregated alert (pre, post, tmstmp)**, in which pre (precondition) and post (postcondition) are described in predicate with readable information. These predicate can be extended to include general attributes of source/ target IP address, source/ target such as SadmindServic (DstIP), to represent there is Sadmind Service running in the host of DstIP. Timestamp is used to combine temporal relationship between aggregated alerts over a time window (such as Δt seconds). In a specific time series, corresponding algorithms are designed to complement aggregated alerts before correlation. For example, if SrcIP = DstIP then **land** attack is characterized in attack type. And in a short period, there is a large traffic all in a sudden from a great number of Source to a specific Targe, then maybe **DDoS** attack is happened.

A general algorithm is given as follows (Fig.1). All fields are denoted as an ordered tuple with four elements, the wildcards are given for default items. For example, ping alert **ICMP PING NMAP** can be denoted as **Probe_ICMP PING(*, livehost (DstIP), 2007-10-23 15:30:20)**.



```
Correlation (A, S)
INPUT: A, a set of n aggregated alerts
OUTPUT: S, a sequence of attack scenario
    Sort aggregated alerts in A such that A[1].tmstmp≤…≤A[n].tmstmp
    sequence S=Φ
    for i=1 to n
        for j=i+1 to n
          if A[i].post=A[j].pre
          then s[i]=A[i] U {A[j]}; break
    return S
```

Fig.1 Causal correlation algorithm

This algorithm can be used to correlate most of the attacks, but not all attacks. For a small number of special attacks, specific algorithms can be designed to complement the correlation mechanism.

C. Reasoning Model for Scenario Recognition

A common characteristic for IDS alerts is that each low-level alert that corresponds to a single attack step (probe, buffer overflow, or other event). The process of connecting the step, that is, correlating alerts from different sensors regarding same or different events and recognizing complex attack scenario is typically manual and slow. Therefore, it would be highly desirable to automate correlate successful alerts and recognize multistage attack scenarios.

Let us consider the following multistep attack scenario as an example: Firstly, an attacker tries to gather sensitive information about the victim host. The information ranges from learning what machines are running in the network to probing for specific services running on the host machines. And **FINGER search query** alert is generated when an attempt is made to query the finger daemon to ascertain the list of some accounts existing on the victim system as a prelude to further comprise. Then **FINGER root query** alert is triggered when an attempt to access information about the administrative account root on a UNIX system is made via the finger service. Finally, **FINGER remote command execution attempt** is generated when a remote command execution exploit against a finger daemon is attempted. Realizing the general attack pattern described in the above, Fig.2 demonstrates the proposed reasoning model for scenario recognition, benefiting from a reliable correlation mechanism for pattern matching.

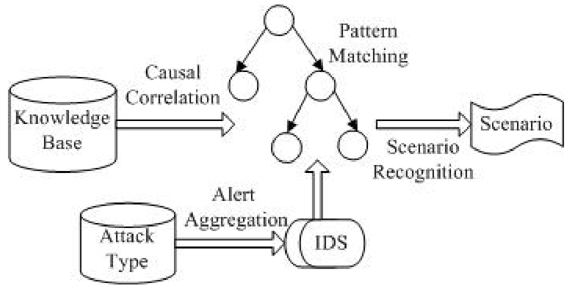


Fig.2. Reasoning model for scenario recognition

As shown in Fig.2, after attack-attribute alert aggregation process, high-level aggregated alerts have been identified with statistics analysis, and then based on causal correlation with time series, attack scenario is constructed using pattern matching.

V. Implementation

In order to further evaluate our reasoning model in practice, we implement the components in an experimental environment (including an attacker machine, a target machine and a machine with Snort sensor, RealSecure and our prototype system).

With regards to typical datasets, we have performed 2000 DARPA intrusion detection scenario-specific datasets [MIT Lincoln Lab 2000] [7]. These experiments were aimed at evaluating the effectiveness of the proposed correlation mechanism in constructing attack scenarios. In the experiment, we replay the datasets in dump format and capture the traffic flow in an isolated network monitored by Snort and RealSecure. There are 4676 alerts from Snort and 922 ones from RealSecure together. Through alert abstraction described in Table 1, these low-level alerts are aggregated into high-level ones and the number of alerts from heterogeneous IDSs is reduced to 322. Then applying the correlation algorithm based on time series and statistical analysis, attack scenario is recognized through the reasoning model.

Fig.3. demonstrates the attack scenario with 2000 DARPA intrusion datasets using our proposed model.

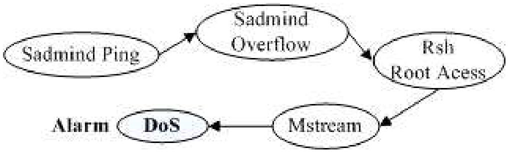


Fig.3. Attack scenario with 2000 DARPA intrusion datasets

With regarding to DDoS attack, Snort does not report the alerts related to communication of the DDoS trojans on the compromised hosts and also the final step of the attack scenario DDoS attack. Specifically, by combing the alerts from Snort and RealSecure, we are able to correctly correlate the five steps of the DDoS scenario starting from probing machines for sadmind service using the *Sadmind Ping* followed by *Sadmind Overflow* and *Rsh Root Access* to gain access to the victim machine. Then the attacker installs *Mstream* DDoS master and agents when break-in succeeds. And then *DoS* attack is launched distributely.

VI. Conclusion and Future Work

In this paper, we developed a method to recognize attack scenario from heterogeneous intrusion detection systems. By using general attack pattern of attack classification, we combine casual correlation mechanism based on time series and statistical analysis to recognize attack scenarios. One issue to be addressed is the finer granularity of multiple alerts from different IDS systems, which benefit the process of attack-attribute aggregation.

Our approach differs from prior work in that it focuses on attack type in alert aggregation. Instead of only depending on the prior knowledge of pre/post-conditions, we correlate the aggregated alerts based on time series and statistical analysis to construct attack scenarios.

Applying collaborative attack modeling to alert aggregation, it does benefit alert correlation to model attack scenario. Through the experimental results with DARPA Data Sets 2000, it demonstrates the potential of the proposed techniques.

Then our future work is to further design the general attack patterns base on collaborative attack modeling, and most importantly, to further improve on modeling attack scenario in real network environment.

REFERENCES

[1] J. Allen, A. Christie, W. Fithen, J. McHugh, J. Pickel and E. Stoner, “State of the Practice of Intrusion Detection Technologies,” Software Engineering Institute of Carnegie Mellon University, PA, USA, Technique Report, January 2000.

[2] C. Kruegel, W. Robertson, and G. Vigna, “Using Alert Verification to Identify Successful Intrusion Attempts,” In Practice in Information Processing and Communication (PIK), vol. 27, no. 4, pp. 219–227, October/December 2004.

[3] Snort. Available: http://www.sourcefire.com/snort.

[4] G. Helmer, J. Wong, M. Slagell, V. Honavar and L. Miller, “A Software Fault Tree Approach to Requirements Analysis of an Intrusion Detection System,” In Proceedings of the 1st Symposium on Requirements Engineering for Information Security, October 2000.

[5] B. Morin, L. M´e, H. Debar, and M. Ducass´e, “M2D2: A Formal Data Model for IDS Alert Correlation,” In Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection (RAID 2002), LNCS 2516, Zurich, Switzerland, pp. 115-137, October 2002.

[6] Zh. Jingmin, M. Heckman, B. Reynolds, A. Carlson and M. Bishop, “Modeling Network Intrusion Detection Alerts for Correlation,” ACM Transactions on Information and System Security, NY, USA, February 2007.

[7] M.I.T Lincoln Laboratory, “2000 DARPA Intrusion Detection Scenario Specific Datasets,” Available: http://www.ll.mit.edu/IST/ideval/data/2000/2000_data_index.htm

[8] B. Schneier, “Attack Trees,” Dr. Dobb's Journal, vol. 24, no. 12, pp. 21-29, December 1999.

[9] J. Mahalati, “Facilitating Alert Correlation Using Resource Trees,” [Master. Thesis], North Carolina State University, 2005.

[10] S. J. Stolfo, W. Fan and W. Lee, “Cost-Based modeling and evaluation for data mining with application to fraud and intrusion detection: Results from the JAM project,” In Proceedings of the DARPA Information Survivability Conference, 2000.

[11] S. Cheung, U. Lindqvist, and M. W. Fong, “Modeling multistep cyber attacks for scenario recognition,” In Proceedings of the DARPA Information Survivability Conference and Exposition, CA: IEEE Computer Society Press, Los Alamitos, pp. 284-292, 2003.

[12] P. Ning, Y. Cui, D. S. Reeves and D. B. Xu, “Techniques and Tools for Analyzing Intrusion Alerts,” ACM Transactions on Information and System Security (TISSEC), vol. 7, no. 2, 2004.

[13] P. Ning, Y. Cui, and D. S Reeves. “Constructing attack scenarios through correlation of intrusion alerts,” In Proceedings of the 9th ACM Conference on Computer and Communications Security, Washington, D.C., pp 245 – 254, November 2002.

[14] P. Ning, D. Xu, C. Healey, and R. St. Amant, “Building attack scenarios through integration of complementary alert correlation methods,” In Proceedings of the 11th Annual Network and Distributed System Security Symposium (NDSS 2004), pp. 97-111, February 2004.

[15] F. Cuppens and R.Ortalo, “LAMBDA: A Language to Model a Database for Detection of Attacks,” In Proceedings of the 3rd International Workshop on the Recent Advances in Intrusion Detection (RAID'2000), Toulouse, France, 2000.

[16] F. Cuppens and A. Miege, “Alert Correlation in a Cooperative Intrusion Detection Framework,” In Proceedings of the 2002 IEEE symposium on security and privacy, 2002.

[17] H. Debar and A. Wespi, “Aggregation and Correlation of Intrusion-detection Alerts,” In Proceedings of the 5th International Symposium on Recent Advances in Intrusion Detection (RAID), 2001.

[18] F. Valeur, G. Vigna, K. Christopher and Richard A. Kemmerer, “A Comprehensive Approach to Intrusion Detection Alert Correlation. IEEE Transactions on Dependable and Secure Computing,” vol. 1, no. 3, 2004.

[19] J. Steffan and M. Schumacher, “Collaborative Attack Modeling,” 17th ACM Symposium on Applied Computing (SAC 2002), Special Track on Computer Security, Madrid, Spain, March 2002.

[20] X. Qin and W. Lee, “Statistical Causality Analysis of INFOSEC Alert Data,” In Proceedings of the 6th symposium on Recent Advances in Intrusion Detection (RAID 2003), Lecture Notes in Computer Science, vol. 2820, pp. 73-93, 2003.

[21] A. Valdes and K. Skinner, “Probablistic alert correlation,” In Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection(RAID), October 2001.

[22] F. Autrel and F.Cuppens, “Using an Intrusion Detection Alert Similarity Operator to Aggregate and Fuse Alerts,” In Proceedings of the 4th Conference on Security and Network, 2005.