데이터 마이닝을 적용한 에이전트 침입 탐지 시스템 설계

정종근*, 이성태, 김용호, 이윤배

*조선대학교 전자계산학과

Design of agent intrusion detection system applying data mining

Jeong Jong Kun*, Lee Sung Tae, Kim Yong Ho, Lee Yun Bae

Dept of computer Science, chosun university

E-mail : jkcom@hanmail.net

요 약

침입 탐지 시스템은 침입 판정과 감사 자료(audit data) 수집 분야에서 많은 연구가 진행되고 있다. 침입 판정은 주어진 일련의 행위들이 침입인지 아닌지를 정확히 판정해야하고 감사 자료 수집에서는 침입 판정에 필요한 자료만을 정확히 수집하는 능력이 필요하다. 최근에 이러한 문제점을 해결하기 위해 규칙기반 시스템과 신경망등의 인공지능적인 방법들이 도입되고 있다. 하지만 이러한 방법들은 단일 호스트 구조로 되어있거나 새로운 침입 패턴이 발생했을 때 탐지하지 못하는 단점이 존재한다. 본 논문에서는 분산된 이기종 간의 호스트에서 사용자의 행위를 추출하여 패턴을 검색, 예측할 수 있는 데이터 마이닝을 적용하여 실시간으로 침입을 탐지하는 방법을 제안하고자 한다.

ABSTRACT

As network security is coming up with significant problem after the major Internet sites were hacked nowadays, IDS(Intrusion Detection System) is considered as a next generation security solution for more reliable network and system security rather than firewall. In this paper, we propose the new IDS model which can detect intrusion in different systems as well as which can make real-time detection of intrusion in the expanded distributed environment in host level of drawback of existing IDS. We implement its prototype and verify its validity. We use pattern extraction agent so that we can extract automatically audit file needed in distributed intrusion detection even in other platforms.

Ⅰ. Introduction

Recently, the famous Internet sites such as Yahoo, Amazon, and CNN have been hacked and their services have been stopped. Therefore, a protection of internal information is essential to electronic commerce through Internet, and thus a new system protection mechanism is required. Currently, firewall alone can protect external attack to a certain degree, but internal illegal action cannot be prevented. Active researches have been proceeding on intrusion detection system monitoring in real-time not only external attack but also internal illegal intrusion. Intrusion detection method is difficult to get significant effect for a single host, because most of Internet sites or internal networks is operating in a distributed environment, not as a single host. In this paper, we propose a model of real-time intrusion detection system based on multiple hosts in a distributed environment using an agent that extracts and analyzes pattern of intruders acting illegally in a system.

Ⅱ. Technical classification of intrusion detection system

Ⅱ-1 Technical analysis of intrusion detection system

Implementation method of intrusion detection system can be classified into three kinds of it as follows:

- A technique of real-time monitoring and analysis of intrusion
- A technique of real-time collection and analysis of packets
- A technique of analysis by audit data

analysis

A technique of real-time monitoring of intrusion detects arbitrary access or modification of unauthorized files, or change of login program. An effective method for real-time intrusion detection is that illegal behaviors occurring from various systems and devices should be monitored real-rime and appropriate measures should be taken. Most of behavior monitoring use audit data provided by operating system. On the other hand, for many-sided detection, they should be used audit data generated by the activation of Webserver, Router, Firewall, and TCP/UDP port. For real-time monitoring of intrusion, system damage can be reduced by taking immediate measures in detection of those behaviors, because intruders usually try to get an authorization of an administrator. A technique of real-time intrusion detection can be divided into a single host-based intrusion detection and multiple hosts-based intrusion one. The former is not suitable for currently multi-platform environment, because it works for a single system only[8]. Multiple hosts-based technique detects intrusion by collecting and analyzing audit data in a distributed environment with recognition of whole network and system as an agent. The role of an agent is to collect and extract audit data with it installed at multiple hosts connecting to network.

III. Analysis of real-time intrusion detection system

Most of intrusion detection systems studied so far is classified by intrusion detection model base and data source base, and they use rule-based detection method. Those systems have a serious problem degrading the performance of whole system due to defects of a process as well as system load, because they perform the detection by a single intrusion detection process. A solution to the problem is to get multiple agents to perform system monitoring, data collection, and intrusion detection in a whole of distributed systems[1]. Those systems cannot also cope flexibly with a new type of attack or a change of system environment, because they do not have self learning function.

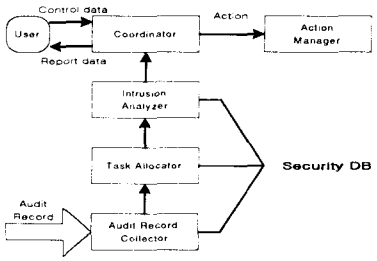

Fig1. Frame of real time IDS

IV. Design of real time data mining intrusion detection system using agent

IV- I Proposed system structure

An agent is a system most suitable for monitoring state of network or system in a distributed environment. Especially, intrusion detection system using an agent is ideal for real-time intrusion detection system in which learning about intrusion information should be made automatically. Fig2 shows the structure of automatic pattern extracting agent. In this paper, we propose automatic pattern extracting agent not only for learning of past intrusion types but also for recognition and learning of new intrusion types[2]. As in Fig2, agent's structure consists of 4 parts of interface agent, pattern extracting agent, profile collecting agent, and process monitoring agent.

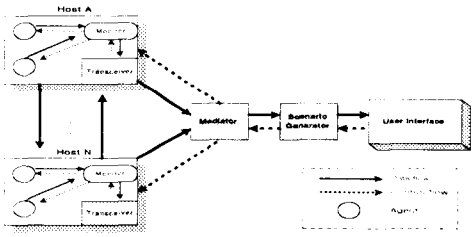

Fig2. Suggested IDS model

Interface agent transmits detection scenario produced from intrusion detection server, or establishes environment fit for target hosts.

Pattern extracting agent extracts audit data only needed for scenario of intrusion detection server out of audit data collected from profile collecting agent. The collected audit data are retransmitted to intrusion detection server, and they are stored in database in case the collected pattern is new. Profile collecting agent and process monitoring agent collects from the

kernel audit data of events occurring from target hosts such as CPU-used time, login-failed ID, and particular port access attempts. Especially, interface agent, which receives scenario of intrusion detection system, orders pattern extracting agent to collect information about profile and process of current users.[3] The following is a class structure algorithm of pattern extracting agent.

Class Patt_extractor_agen
rcv_scenario();
request_profile();
rcvprofile();
request_process();
Clsaa store_pattern_DB
request_pattern();
stroe_DB();
Class InterF_agen
request_pattern();
rcv_pattern();
sendpattern();

In case target systems are different from each other, a problem comes up with a format of audit file. In this paper, we selected a standardized mode of extracted audit file for a solution to the problem. Audit files which are moved to pattern extracting agent are regenerated into a standardized format at Log Generator.

Ⅳ-Ⅱ. Log audit data standardization

Audit data techniques of intrusion detection system studied so far, which have system-dependent features, do not fit to support different environment. In this paper, log data analyzer maintains a consistent log audit data structure with audit data standardization, using log filter, of log data analyzed after collected from each scenario.[4] Its generation structure is shown in Fig3. First, log information needed at log analyzer of each operating system is collected. Only log fields needed by log processor are extracted through log filter. And then, log processor transforms them into a standardized format. It genetates audit data needed for intrusion detection system in a standardized structure.

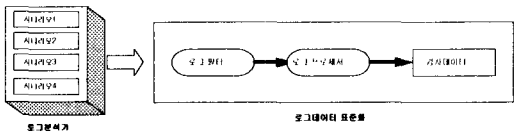

Fig3. Structure of log processor

Ⅴ. Performance evaluation of a proposed system

This paper used in experimental materials 4 scenarios generated from scenario generator and their related commands. Those materials corresponds to the commands, files, or directories used in UNIX, and a threshold value was given for intrusion decision. If it is too high, overall detection rate is low, and a critical error of recognizing an intrusion as a normal use can occur, even though a correctness of intrusion decision is high. If it is too low, overall detection rate is high, while a detection correctness is low. Therefore, this paper got an intrusion decision engine to adjust a threshold value. Thus, an agent collects user's log file as indicated in scenarios, and uses it for an intrusion decision. This paper gets a threshold value to be adjusted by a system administrator according to user's number or system throughput

Fig4. represents a detection rate in graph according to an adjustment of a threshold value for an intrusion decision.

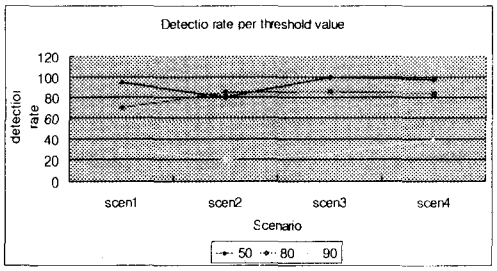

Fig4. Intrusion detection ratio variation following by critical value

Fig4. represents that lower threshold value indicates higher intrusion detection rate and shorter consuming time, but higher threshold does lower detection rate, higher correctness, and longer detection time. A feature of this system is that it extracts audit data needed for an intrusion detection at agent's step and transforms them into a standardized format, which minimizes a job load in an intrusion

detection host, and thus it can cope appropriately with system situations by making possible an adjustment of a threshold value suitable for a detection[2].

Ⅵ. Conclusion

Intrusion detection system of pattern extracting agent proposed in this paper optimized an efficiency of intrusion detection between different systems especially through a standardization phase of audit data after real-time collection of data needed for intrusion detection at an agent in distributed hosts. It makes out intrusion scenarios at scenario generator in order to collect data needed for intrusion detection at an agent, and then it transmits them to an agent. An agent collects necessary data only through its immediate response. In a standardization phase of audit data, a load of intrusion detection system deteriorates overall system speed, in case multi-intrusions occur in multiple hosts. Therefore, it minimized a system load in intrusion decision by drawing up in a single format audit data collected in different systems with a standard audit file format. However, an automatized scenario update and intrusion techniques need to continue to be studied in a detection of hacker's abnormal intrusion, because intrusion techniques are getting diverse.

reference

[1] S.Kumar and E.Spafford, "A pattern matching model for misuse intrusion detection." Seventeenth National Computer Security Conference, Baltimore, MD, October 1994, 11-21.

[2] T. lane and C. E. Brodley. "Detecting the abnormal: Machine learning in computer security", Technical Report TR-ECE 97-1, Prudue University, West Lafayette, IN, 1997.

[3] Neil Crowe and Sandra Schiavo, "An Intelligent Tutor for Intrusion Detection on Computer System", code Cs/rp, Department of Computer Science, Naval postgraduate school monterey, 1997

[4] Crosbie M, Spafford E, "Defending a Computer System using Autonomous Agents," Technical Report, Purdue University, Department of Computer Science, 1996.

[5] 이윤배 외 4인, "데이터 마이닝 기법을 적용한 최적 침입 탐지 모듈 설계", 1999 춘계 정보과학회 논문집

[6] "정보시스템 침해사고 방지기술 개발에 관한 연구", 정보보호센터, 1999.1