LIDS: Learning Intrusion Detection System

M. Dass **J. Cannady** **W.D. Potter**

Artificial Intelligence Center,
University of Georgia,
Athens, Georgia,
U.S.A.
dass@uga.edu; j.cannady@computer.org; potter@uga.edu

Abstract

The detection of attacks against computer networks is becoming a harder problem to solve in the field of network security. The dexterity of the attackers, the developing technologies and the enormous growth of internet traffic have made it difficult for any existing intrusion detection system to offer a reliable service. However, a close examination of the problem shows that there usually exists a behavioral pattern in the attacks that can be learned and can be used to detect intrusions more effectively. Thus, there is a requirement for a system with learning and adapting capabilities for optimal performance. This paper discusses the design of a Learning Intrusion Detection System (LIDS) that includes a blackboard-based architecture with autonomous agents. It has the capability for online learning, which may result in better performance than present systems. This feature enables the system to adapt to changes in the network environment as it assimilates more network data.

Keywords: Intrusion Detection, Blackboard Architecture, Autonomous Agents, Artificial Neural Networks.

Introduction

With the rapid increase in vulnerable Internet applications and automated attack scripts, intrusions of networked systems have become an increasing problem in the field of information technology. Every year, the business industry loses a huge amount of revenue due to data manipulation caused by computer network intruders. As a result, there has been an increasing requirement to effectively protect crucial business information with a reliable, robust and flexible intrusion detection system. There are many commercially available Intrusion Detection Systems (IDS) in the market. Unfortunately they are expensive and of only limited reliability. The increasing complexity of the Internet and the maintenance cost of these systems is a setback to the performance of IDSs. This has led to worldwide research interest in developing the Next Generation Intrusion Detection Systems, which are able to learn and adapt to the network environment for optimal performance. Some of the recent work on developing effective network security highlights new areas of research, which include artificial intelligence [Lane et al. 1999], data mining [Lee et al. 2000], statistical techniques [Denning 1987], agent frameworks including autonomous agents [Balasubramaniyan et al. 1998], intelligent agents [Carver et al. 2000] and mobile agents [Asaka et al. 1999] for distributed intrusion detection [Asaka Taguchi et al. 1999]. However, there has been only a limited amount of research carried out in developing a Learning Network Security System that can become more intelligent while it is detecting intrusions. This paper proposes a blackboard based three-tier autonomous learning agent architecture that has learning and adaptation capability for improved performance.

Intrusions and Priori Approaches in Intrusion Detection System

As the term “Intrusion Detection System” suggests, we are trying to develop a network security system that will detect misuse behavior in the network data stream. These security systems collect network data from the system and audit them in order to detect intrusions. Normally IDSs are located on a centralized server, but some distributed types of IDSs can be placed on different workstations to detect intrusions. The proposed architecture is a server based or centralized IDS.

The process of Intrusion Detection can be defined as the problem of identifying individuals who are using computer network resources without authorization or attempting to prevent authorized users from accessing network resources. In an organization, intrusions can take place from the Internet or from inside the organization’s computer network system. This highlights the two different types of Intrusion Detection Systems; Host Based Intrusion Detection System and Network Based Intrusion Detection System. A Host Based Intrusion Detection System can be defined as a security system that is capable of detecting inside abuses in a computer network. A Network Based

Intrusion Detection System is capable of identifying abusive uses or attempts of unauthorized usage of the computer network from outside the system. This paper describes a Network Based Intrusion Detection System that will use computational intelligence techniques to detect intrusions.

There are several forms of network intrusions:

- Denial-of-service Attack - This is particularly a serious form of attack that has resulted in damages worth millions of dollars over the past few years. While a significant problem, DoS attacks are usually quite simple. They typically involve an attacker disabling or rendering inaccessible a network-based information resource.
- Guessing rlogin Attack – Here the intruder tries to guess the password that protects the computer network in order to gain access to it.
- Scanning Attacks – The intruder goes about scanning different ports of the victim's system to find some vulnerable points from where they can launch other attacks.

Most of the present approaches in detecting intrusions utilize some form of rule-based analysis. Rule-Based analysis relies on predefined rule-sets that are provided by an administrator, automatically created by the system, or both. Expert Systems are the most common form of rule-based intrusion detection approaches [Denning 1987]. Rule-based systems suffer from the inability to detect attack scenarios that may occur over an extended period of time. They also lack flexibility in the rule-to-audit record representation [Cannady 1998]. Slight variations in the attack sequence may reduce the effectiveness of the system.

An increasing amount of research has been conducted on the usability of neural networks in accurately detecting network attacks [Cannady 1998], and efforts have been made to integrate a rule-based system with a neural network to develop a high performance Intrusion Detection System. Research efforts have also been made to use non-traditional AI-based techniques like Genetic Algorithms [Ludovic 1998], Data Mining [Lee et al. 2000] and Pattern Recognition Techniques [Denning 1987] to develop a high performance IDS. Nonetheless, little effort has been applied to the development of an approach that possesses the capability for continuous learning. Researchers have primarily tried to identify different innovative techniques of detecting intrusions, but they have usually overlooked the potential of a learning system that can adapt itself in the network environment and give high performance with increased experience.

Blackboard and Autonomous Agents

The blackboard architecture is considered as one of the most general and flexible knowledge system architectures for building decision-based applications. It is highly preferred over other alternatives due to its modularity, dynamic control, generality, concurrency, high design efficiency, robustness and ability in dealing with multiple knowledge sources. As a result, the blackboard-based architecture is considered to be a good solution in developing LIDS.

The proposed architecture will include the use of Autonomous Agents. For the proposed architecture, we implement software agents that perform certain security monitoring functions at a host. The agents are independently running entities whose performance is not affected by any other entities. These kinds of agents are very useful in network security because they run continuously, can resist subversion and have minimal overhead. They are also configurable, easily adaptable, scalable, dynamically reconfigurable and degrade gracefully. The proposed LIDS architecture consists of autonomous agents that are integrated in a blackboard-based architecture.

Proposed Architecture

The use of blackboard techniques and autonomous agents [Balasubramaniyan et al. 1998] in detecting network intrusions is not a new concept [Balasubramaniyan et al. 1998], [Dasgupta et al. 2002]. Dasgupta described how blackboard-based agent architecture helps in detecting intrusions [Dasgupta et al. 2002]. Dasgupta also developed a distributed blackboard architecture that is embedded among the agents. A manager agent controls the monitoring, decision and action agents. The unidirectional flow of information in the system has a major impact on the flexibility of the system. Balasubramaniyan applied rule-based autonomous agents to detect intrusions and as a result faced the same problems as faced by rule-based Intrusion Detection Systems. Rule-based systems lack the flexibility to identify new attacks in the network data stream and must be updated frequently to remain current with the evolving threat posed by network attackers. LIDS have a blackboard-based autonomous agent architecture that is designed in a multi-tier format (Figure 1).

There are eight autonomous agents in the system that interacts with the blackboard to perform their actions. Generally, a blackboard system consists of three components; they are the action agents, a blackboard and a control mechanism that will guide these agents [Penny 1989]. In the proposed system, there is no control/manager agent, but there is a control pattern embedded in each agent that guides their activities.

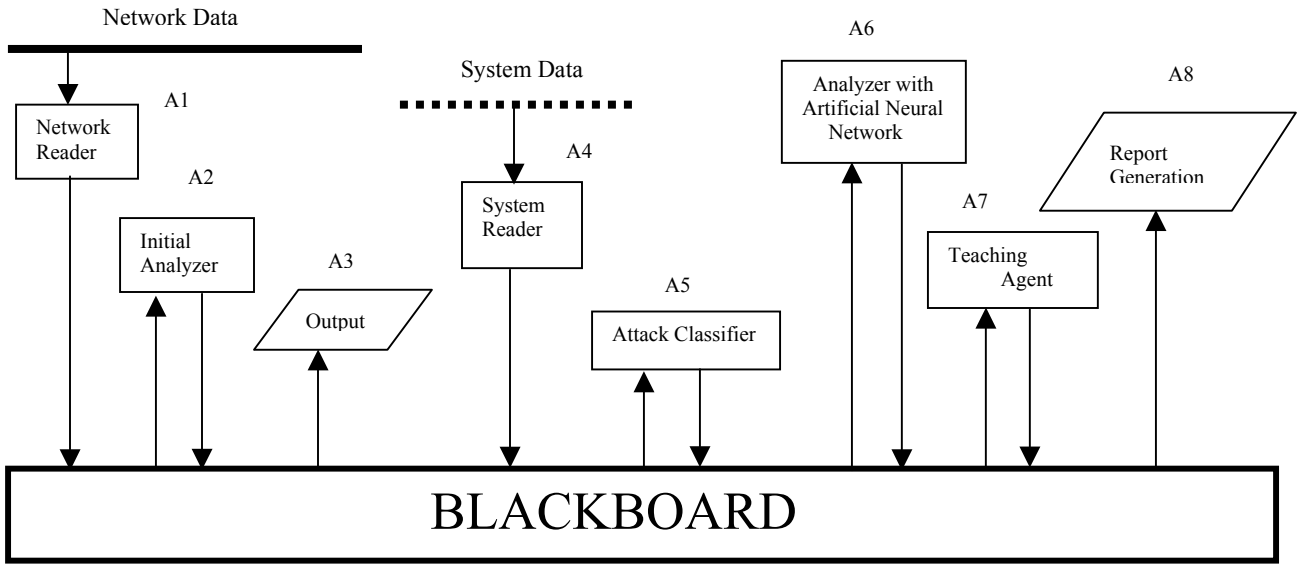

Figure 1: LIDS Architecture

The first agent (A1) is called the Network Reader. It collects network data with the help of a program called tcpdump. Tcpdump is a network utility tool that records network data in a specific format. The A1 autonomous agent collects network data in groups of 1000 data packets (network activity information) and pastes them on the blackboard. The second agent (A2) is called the Initial Analyzer. It continuously observes the blackboard and whenever it finds any data that need to be analyzed, as the data posted by the first agent, it performs its action. It consists of a Rule-based classifier analyzer as a PROLOG predicate. This analyzer reports to the blackboard whether the data set is clean or not. If some trace of probable attack is found in the data set, it also suggests the type of attack. The third autonomous agent (A3) is the output agent. It helps in displaying any early alert to the security auditors or the administrator of the system, like the information posted by the analyzer agent (A2).

The fourth autonomous agent (A4) is the system reader. This agent gathers system specific information of the protected system and posts it on the blackboard. These system data are very helpful in detecting the extent of damage caused by any attack. The type of information gathered includes Available Network Bandwidth, CPU Usage, Network Packets/second, Memory usage, Number of connections, Connection attempts, Protocol, Source address to destination ports ratio (variety of ports accessed) and Packet length.

There are many sub-class attacks that fall under one kind of attack. For example, a denial-of-service attack can be separated into Ping Flood Attack, a UDP Packet Storm Attack, an FTP Brute Force Attack and so on. The fifth agent (A5) is the attack classifier that identifies different sub-classes of intrusions present in the network data. This agent sends the system information from the blackboard to a micro genetic algorithm based classifier that uses the multiple-fault diagnosis concept to perform the above function and posts its result back to the blackboard. The result states which kind of attack is present and what is its probability of presence in the dataset.

The ultimate purpose of an Intrusion Detection System is to identify the affected network data with some degree of confidence. This is achieved by the next autonomous agent (A6) or the main analyzer. This agent consists of a set of different kinds of Artificial Neural Networks (ANN). It looks for the different kinds of intrusions present in the dataset from the information posted on the blackboard and decides which ANN is suitable for its analysis. If no attack is present in the data set, it flags the result.

Another objective of LIDS is to learn about new attacks while actively engaged in the detection process. This is achieved by the seventh agent (A7) or the teaching agent. The initial analyzer (A2) is powered by a rule-based classifier system. This type of classifier system has a rule set in it, which is in the form of facts. As mentioned above, the analyzing agent (A2) audits the network data recorded by the network reading agent (A1) and reports whether the data has intrusion in it or it is clean data. If the analyzer (A2) finds a new network pattern and reports an intrusion alert, and in later process, it is found that the network-data is clean, the teaching agent (A7) will update the rule-set of analyzer (A2). Therefore, whenever an initial analysis has

resulted in a false alarm or whenever a new type of intrusion is detected that has no supporting rules in the rule-set, we update the rules with the help of the teaching agent.

The final autonomous agent is the Report Generation agent. It generates reports for the system administrator based on the information posted on the blackboard. As mentioned earlier, one of the components of the blackboard architecture is the control mechanism. Since the proposed architecture is autonomous-agent based, there is no agent manager and hence there is a problem in implementing autonomous behavior with sequential processors. The architecture proposed here has a control pattern embedded in the agents. This pattern allows the last agent to look at the blackboard first and the first agent last in order to ensure that each agent gets a chance at least once to look at the blackboard in one process cycle.

The proposed architecture is capable of handling most of the problems faced by the present approaches. It has an online learning mechanism that updates the rules of the analyzer and uses Artificial Neural Networks to analyze the data. These features handle the problems faced by [Denning 1987] where the rule updating and lack of flexibility in the rule-to-audit record representation was a problem. In most of the other approaches using autonomous agents such as [Balasubramaniyan et al. 1998] [Dasgupta et al. 2002], there lacks a common data pool. A common data pool like our blackboard is very important when dealing with various analyzing techniques. Detailed information about intrusions is very important for security officials. A common data pool will also help in storing audited data for future reference. Moreover, this work is unique as this architecture represents a common platform for all the different analysis techniques. Adding a data mining technique or a genetic algorithm as an analyzer to the system will be very easy to implement.

We have completed developing Network Reader Agent (A1), Initial Analyzer Agent (A2), Attack Classifier Agent (A5) and ANN Analyzer Agent (A6). We are using readily available DSSTools to create the blackboard environment. The Attack Classifier as discussed above is a Micro Genetic Algorithm (GA). This GA along with the ANN is in the form of dlls written in C++. They are called by the main agents that are written in PROLOG.

Conclusion and Future Direction

In designing the proposed architecture we have tried to use the best AI techniques for the different knowledge–based problems. This hybrid approach should fulfill the deficiencies of other systems and its learning capability can also make it more efficient in a dynamic network environment. The flexibility of the blackboard architecture will allow us to add more features in the future.

All of these agents are to be written in PROLOG. We will be using DSS tools [Zhu 1995] for developing the blackboard system. Some of these agents perform analysis of the network data with the help of Artificial Neural Networks (ANN), written as dlls in C++. We have already developed the C++ dll for the analyzers. The only part remaining is to develop the remaining agents and to integrate all of them in the system. The DSS tool that will form the blackboard environment is readily available.

References

Lane, T., and Brodley, C. E. 1999. Temporal Sequence Learning and Data Reduction for Anomaly Detection. In the *Proceedings of ACM Transaction on Information and System Security*, Vol. 2, No. 3, August 1999.

Lee, W., Stolfo, S., and Mok, K. 2000. Adaptive Intrusion Detection: A Data Mining Approach. In Artificial Intelligence Review, Kluwer Academic Publishers, 14(6): 533 – 567, December 2000.

Denning, D. E. 1987. An Intrusion-Detection Model. In *IEEE Transaction on Software Engineering*, Vol. Se-13, No. 2, February 1987, 222-232.

Asaka, M., Taguchi, A., and Goto, S. 1999. The implementation of IDA: An intrusion detection agent system. In *Proceedings of the 11th FIRST Conference*, June 1999.

Balasubramaniyan, J., Fernandez, J. O., Isacoff, D., Spafford, E., and Zamboni, D. 1998. An Architecture for Intrusion Detection Using Autonomous Agents. In COAST Technical Report 98/5, Purdue University, June 1998.

Carver, C. A., Hill, J. M. D., Sudru, J. R., and Pooch, U. W. 2000. A Methodology for Using Intelligent Agents to Provide Automated Intrusion Response. In *IEEE Systems, Man and Cybernetics Information Assurance and Security Workshop*, West Point, N.Y., June 2000.

Asaka, M., Okazawa, S., Taguchi, A., and Goto, S. 1999. A Method of Tracing Intruders by Use of Mobile Agents. In *INET' 99*, June 1999

Cannady, J. 1998. Artificial Neural Networks for Misuse Detection. In *Proceedings of Recent Advances In Intrusion Detection* 1998.

Ludovic, M.E. 1998. GASSATA, a Genetic Algorithm as an Alternative Tool for Security Trails Analysis. In *Proceedings of Recent Advances In Intrusion Detection* 1998

Dasgupta, D., Gonzalez, F., Yallapu, K., Gomez, J., Yarramsettii, R., Dunlap, G. and Greveas, M. 2002. CIDS: An Agent-based Intrusion Detection System. In CS Technical Report No. CS-02-001., Feb, 2002.

Penny, Ni. H. 1989. BlackBoard Systems. In Handbook of Artificial Intelligence, Vol IV, edited by Avron Barr, Paul R. Cohen and Edward A. Feigenbaum. Reading, MA: Addison-Wesley Publishing Company Inc.

Zhu, G. 1995. DSSTOOLS : A toolkit for development of Decision Support Systems in PROLOG. M.S. thesis, AI Center, University of Georgia.