

Novel Attack Detection Using Fuzzy Logic and Data Mining

Norbik Bashah Idris Bharanidharan Shanmugam
Centre for Advanced Software Engineering
Universiti Teknologi Malaysia – City Campus
Jalan Semarak, Kuala Lumpur 54100 MALAYSIA

Abstract: - Intrusion Detection Systems are increasingly a key part of systems defense. Various approaches to Intrusion Detection are currently being used, but they are relatively ineffective. Artificial Intelligence plays a driving role in security services. This paper proposes a dynamic Intelligent Intrusion Detection System model, based on specific AI approach for intrusion detection. The technique that is being investigated includes fuzzy logic with network profiling, which uses simple data mining techniques to process the network data. The proposed hybrid system combines anomaly and misuse detection. Simple fuzzy rules, allow us to construct if-then rules that reflect common ways of describing security attacks. Suspicious intrusions can be traced back to its original source and any traffic from that particular source will be redirected back to them in future. Both network traffic and system audit data are used as inputs for the experimental needs.

1.0 Introduction.

Information has become an organization's most precious asset. Organizations have become increasingly dependent on information, since more information is being stored and processed on network-based systems. A significant challenge in providing an effective and efficient protective mechanism to a network is the ability to detect novel attacks or any intrusion works and implement counter measures. Intrusion detection is a critical component in securing information systems. Intrusion detection is implemented by an Intrusion detection system. Today, we can find many commercial Intrusion Detection Systems available in the market, but they are restricted in their monitoring functionality and they need frequent updates and patches. The wide spread use of e-commerce, has increased the necessity of protecting the system to a very high extend. Confidentiality, Integrity and Availability of information are major concerns in the development and exploitation of network based computer systems. Intrusion Detection System, can detect, prevent and react to the attacks. Intrusion Detection has become an integral part of the information security process. However, it is not technically feasible to build a system with no vulnerabilities; as such, intrusion detection continues to be an important area of research.

The remaining part of this paper is organized as follows: Section 2 gives an overview of current Intrusion Detection Systems and also on the usage

of fuzzy and data mining techniques. Section 3 elucidates the overview of our proposed architecture. Section 4 briefs about the Traceback framework and Section 5 summarizes the work and points out what we will be doing in future.

2.0 Intrusion Detection Systems.

2.1 An Overview of the Current Intrusion Detection Systems

Intrusion Detection is defined [1] as the process of intelligently monitoring the events occurring in a computer system or network and analyzing them for signs of violations of the security policy. The primary aim of Intrusion Detection Systems (IDS) is to protect the availability, confidentiality and integrity of critical networked information systems. Intrusion Detection Systems (IDS) are defined by both the method used to detect attacks and the placement of the IDS on the network. IDS may perform either misuse detection or anomaly detection and may be deployed as either a network-based system or a host-based system. This results in four general groups: misuse-host, misuse-network, anomaly-host and anomaly-network. Misuse detection relies on matching known patterns of hostile activity against databases of past attacks. They are highly effective at identifying known attack and vulnerabilities, but rather poor in identifying new security threats. Anomaly detection will search for something rare or unusual by applying statistical measures or artificial intelligence

methods to compare current activity against historical knowledge. Common problems with anomaly-based systems are that, they often require extensive training data for artificial learning algorithms, and they tend to be computationally expensive, because several metrics are often maintained, and need to be updated against every systems activity. Some IDS combine qualities from all these categories (usually implementing both misuse and anomaly detection) and are known as hybrid systems. Artificial Intelligence techniques have been applied both to misuse detection and also for anomaly detection. SRI's Intrusion Detection Expert System (IDES) [2] encodes an expert's knowledge of known patterns of attack and system vulnerabilities as if-then rules. Time-based Inductive Machine (TIM) [3] for intrusion detection learns

sequential patterns. Recently, techniques from data mining area have been used to mine normal patterns from audit data [4,5,6]. Several approaches applying artificial neural networks in the intrusion detection system have been proposed [7,8,9]. NeGPAIM [10] is based on trend analysis, fuzzy logic and neural networks to minimize and control intrusions. Existing intrusion detection systems especially commercial intrusion detection systems that must resist intrusion attacks are based on misuse detection approach, which means these systems will only be able to detect known attack types and in most cases they tend to be ineffective due to various reasons like non-availability of attack patterns, time consumption for developing new attack patterns, insufficient attack data etc.,

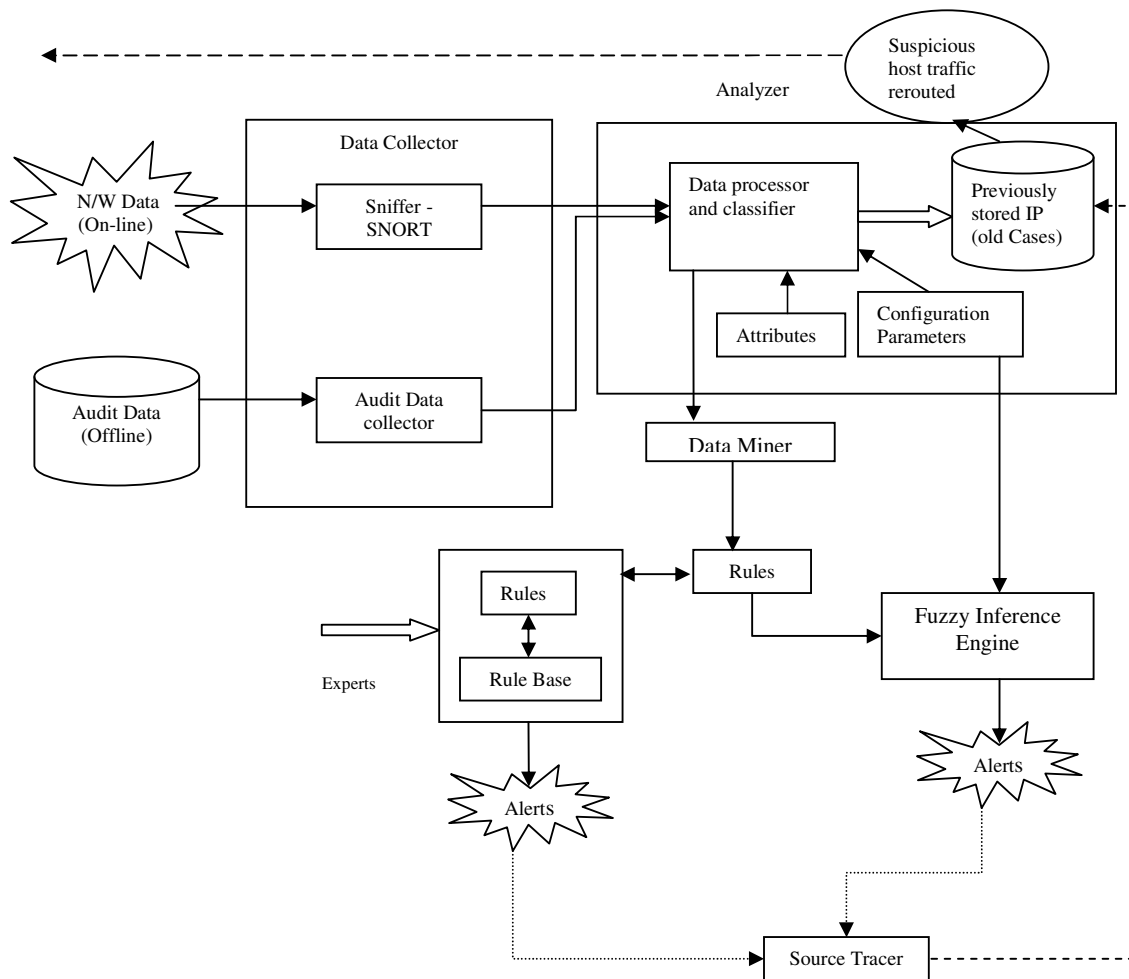


Fig 1. Intelligent Intrusion Detection System Model

2.2 Computer Attack Categories

DARPA [11] categorizes the attacks into five major types based on the goals and actions of the attacker. DoS attacks try to make services provided by or to computer users to be restricted or denied. For example, in SYN-Flood attack, the attacker floods the victim host with more TCP connection requests than it can handle, causing the host to be unable to respond even to valid requests. Probe attacks attempt to get information about an existing computer or network configuration.

Remote-to-local (R2L) attacks are caused by an attacker who only has remote access rights. These attacks occur when the attacker tries to get local access to a computer or network.

User-to-root (U2R) attacks are performed by an attacker who has rights at user level access and tries to obtain super user access.

Data attacks are performed to gain access to some information to which the attacker is not permitted to access.

Many R2L and U2R goals are for accessing the secret files.

2.3 Data Capturing using SNORT

Snort is mainly a Network Intrusion Detection System (NIDS); it is Open Source and available for a variety of unices. Snort also can be used as a sniffer to troubleshoot network problems. Basically there are three modes in which Snort can be configured: sniffer, packet logger, and network intrusion detection system. Sniffer mode simply reads the packets off of the network and displays them in a continuous stream on the console. Packet logger mode logs the packets to the disk. Network intrusion detection mode is the most complex and configurable configuration, allowing Snort to analyze network traffic for matches against a user defined rule set and performs several actions based upon what it sees. We have configured Snort in the packet logger mode for our experimental needs.

2.4 Data Mining and Association Rules

Data Mining is the automated extraction of previously unrealized information from large data sources for the purpose of supporting actions. The rapid development in data mining has made available a wide variety of algorithms, drawn from the field of statistics, pattern recognition, machine learning and databases. Specifically, data mining approaches have been proposed [4,12] and used for

anomaly detection. Association rule algorithms find correlations between features or attributes used to describe a data set. The most popular algorithm for mining rules based on two valued attribute is APRIORI. This algorithm leads to the problem of categorizing numerical attributes. A solution to this problem was given in [13] by transforming quantitative variables into a set of binary variables by partitioning the domain variables into discrete intervals. This approach, however, suffered from "sharp boundary problem". An alternative solution, [14] using fuzzy, offered a smooth transitions from one fuzzy set to another.

2.4 Fuzzy Logic and Intrusion Detection

Applying fuzzy methods for the development of IDS yield some advantages, compared to the classical approach. Therefore, Fuzzy logic techniques have been employed in the computer security field since the early 90's. The fuzzy logic provides some flexibility to the uncertain problem of intrusion detection and allows much greater complexity for IDS. Most of the fuzzy IDS require human experts to determine the fuzzy sets and set of fuzzy rules. These tasks are time consuming. However, if the fuzzy rules are automatically generated, less time would be consumed for building a good intrusion classifier and shortens the development time of building or updating an intrusion classifier. The model suggested in [15] building rare class prediction models for identifying known intrusions and their variations and anomaly/outlier detection schemes for detecting novel attacks whose nature is unknown. The latest in fuzzy is to use the Markov model. As suggested in [16] a Window Markov model is proposed, the next state in the window equal evaluation to be the next state of time t , so they create Fuzzy window Markov model. As discussed in [17] proposes a technique to generate fuzzy classifiers using genetic algorithms that can detect anomalies and some specific intrusions. The main idea is to evolve two rules, one for the normal class and other for the abnormal class using a profile data set with information related to the computer network during the normal behavior and during intrusive (abnormal) behavior. Fuzzy preference relation is another method applied to intrusion detection based on fuzzy satisfaction function. This is applied for comparison of attack signatures. Fuzzy signatures (their gamma resolution sets) are combined by fuzzy operators [18]. The work [19] proposes a dynamic approach that tries to discover known or unknown intrusion patterns which uses Support Vector Machine [20]. A dynamic fuzzy

boundary is developed from labeled data for different levels of security needs.

In this paper we propose a mechanism for IDS which utilizes fuzzy logic along with data mining technique. Our proposed method is the modified version of FIRE [21] system. The FIRE system uses a simple data-mining algorithm to identify the features that will be helpful in detecting attacks. The security administrator uses the fuzzy sets produced by the system to create fuzzy rules. However, here we will propose a mechanism to automate the rule generation process and reduce the human intervention. AI techniques have also been explored to build intrusion detection systems based on knowledge of past behavior and normal use. They have shown potentiality for anomaly detection with limited ability.

3 Goals and Proposed Architecture

Our aim is to design and develop an Intelligent Intrusion Detection System (IIDS) that would be accurate, low in false alarms, not easily cheated by small variations in patterns, adaptive and be of real time. In our model, we use SNORT [22], a leading and famous open source packet sniffer. The data processor and classifier summarizes and tabulates the data into carefully selected categories i.e. the attack types are carefully correlated. This is the stage where a kind of data mining is performed on the collected data. In the next stage, the current data is compared with the historical mined data to create values that reflect how new data differs from the past observed data. The inference engine works based on Mandami inference mechanism. Based on our previous research work we conclude that this mechanism would suit our research requirements. Based on the facts from the analyzer, the decision will be taken whether to activate the detection phase or not. If the detection phase is activated then an alert will be issued and the tracer phase will be initiated. This phase will trace back to the intruders original source address location. We propose a framework for tracing the abnormal packets back to its original source based on single packet. This tends to be the most tedious phase of the project. Once the original path has been identified and verified then all the attacks from that particular host will be redirected to their source in future. SNORT_INLINE [23] has been proven to be the best in changing the appropriate packet values. Figure 1 is the outcome of our initial research work.

3.1 Attributes

Prior to any data analysis, attributes representing relevant features of the input data (packets) must be established. The set of attributes provided to the Data Analyzer is a subset of all possible attributes pertaining to the information contained in packets headers, packet payloads, as well as aggregate information such as statistics in the number and type of packets or established TCP connections. Attributes are represented by names that will be used as linguistic variables [24] by the Data Miner and the Fuzzy Inference Engine.

3.1.1 Data analyzer

Once attributes of relevance have been defined and the data source identified, Data Analyzer is employed to compute configuration parameters that regulate the operation of IDS. This module analyzes packets and computes aggregate information by grouping packets. Packets can be placed in fixed size groups (*s-group*) or in groups of packets captured in a fixed amount of time (*t-group*). Each *s-group* contains the same number of packets covering a variable time range and each *t-group* contains a variable number of packets captured over a fixed period of time.

3.1.2 Rules

Rules are expressed as a logic implication $p \rightarrow q$ where p is called the antecedent of the rule and q is called the consequence of the rule.

3.2 Data Miner

A variation of Kuok's [14] algorithm is used to implement the Data Miner, which allows for efficient, single-pass, record processing by partitioning data into hierarchical files. We plan to integrate the Apriori and Kuok's algorithm which is capable of discovering association rules for binary, categorical and numerical attributes. The final output of the algorithm is a set of rules that meet the confidence and support constraints given as input. These constraints are quantitative qualifiers used to evaluate the relevance of an association rule. *Support* of a rule is a measure of the fraction of the entire data set for which all predicate terms of the rule hold true. *Confidence* of a rule is a measure of the fraction of the data set for which, if the antecedent holds true, then the consequence holds true. For instance, consider a network where 85% of

all the traffic is *TCP* and 50% of all *TCP* traffic is *HTTP*. In that case, the association rule $TCP \rightarrow HTTP$ has a support of 42.5% ($0.85 * 0.50 * 100$) and a confidence of 50%.

4.0 Traceback framework

The design of the IP protocol makes it difficult to reliably identify the originator of an IP packet. Even in the absence of any deliberate attempt to disguise a packet's origin, widespread packet forwarding techniques such as NAT and encapsulation may obscure the packet's true source. Techniques have been developed to determine the source of large packet flows, but, to date, no system has been presented to track individual packets in an efficient and scalable fashion. Unfortunately, the anonymous nature of the IP protocol makes it difficult to accurately identify the true source of an IP datagram if the source wishes to conceal it. The network routing infrastructure is stateless and is based largely on destination addresses; no entity in an IP network is officially responsible for ensuring if the source address is correct. Many routers employ a technique called *ingress filtering* [25] to limit source addresses of IP datagrams from a stub network to addresses belonging to that network, but not all routers have the resources necessary to examine the source address of each incoming packet, and ingress filtering provides no protection on transit networks.

The tracing of the source of the packet is needed for obtaining the exact information about the intruder. The IDS will be providing information that an exceptional event has occurred, the packet and the time of the attack. Once a trace back is requested, a query message consisting of the packet, egress point and the time of receipt is sent to all the Local Data Managers (LDM). Time is critical as this must take place while the appropriate values are still resident at the DC (Data Collector). Once the values are safely transferred to TM, the trace back process will no longer be under real-time constraints. Local Data Manager is responsible for a particular network. Later LDM responds with the partial attack graph and the packet as it entered the region. The attack graph either terminates within the region managed by the LDM, in which a source has been identified, or it contains nodes to the edges of the other LDM network region. Next, TM sends a query to the LDM adjacent of that edge node.

5.0 Summary and Future Work

A hybrid system has been proposed for aiding network personal in the task of computer intrusion detection. We have combined fuzzy logic with data mining to provide efficient technique for both misuse and anomaly based intrusion detection. This model is now at an infant stage of development. Initial results are promising and encouraging. Our long term goal is to make this system implement in a real time environment. More results can be obtained once we finish deploying the system.

6.0 References

- [1] Bace R.G Intrusion Detection, Technical Publishing (ISBN 1-57870-185-6)
- [2] Lunt. T. "Detecting intruders in computer systems". Conference on auditing and computer technology, 1993
- [3] Teng H., Chen K., and Lu S., "Adaptive real time anomaly detection using inductively generated sequential patterns". IEEE computer society symposium on research in security and privacy, California, IEEE Computer Society 278-84 1990
- [4] Lee, Stolfo S., Mok K., "Mining audit data to build data to build intrusion detection models." Fourth international conference on knowledge discovery and data mining, New York, AAAI Press 66-72, 1998
- [5] Mukkamala, R., Gagnon J., Jaiodia S., "Integrating data mining techniques with intrusion detection methods". Research Advances in Database and Information systems security, 33-46, 2000
- [6] Stolfo S., Lee, Chan. "Data mining-based Intrusion detectors : An overview of the Columbia IDS" Project SIGMOD Record Vol 30, No 4, 2001
- [7] Debar, Becker M., Siboni D., "A neural network component for an intrusion detection system." IEEE Computer Society Symposium on Research in Computer Security and Privacy, 240-250 1992
- [8] Tan K., "The Application of Neural Networks to UNIX Computer security". IEEE International conference on Neural Networks Vol 1, 476-481 1995
- [9] Wang J., Wang Z., Dai K., "A Network intrusion detection system based on ANN", InfoSecu04, ACM 2004 (ISBN1-58113-955-1)
- [10] Botha M., Solms R., Perry K., Loubser E., Yamoyany G., "The utilization of Artificial Intelligence in a Hybrid Intrusion Detection System", SAICSIT, 149-155 2002
- [11] MIT Lincoln Laboratory, 1999 DARPA intrusion detection evaluation design and procedure, DARPA Technical report Feb 2001
- [12] Dokas .P, Ertoz L., Kumar V., Lazarevic A., Srivastava J., Tan P., "Data Mining for Network Intrusion Detection" Proceedings of NSF Workshop on Next Generation Data Mining 2002
- [13] Agrawal R., Srikant R., "Fast algorithms for mining association rules" 20th international conference on very large databases September 1994
- [14] Kuok C., Fu A., Wong M., "Mining fuzzy association rules in databases" SIGMOD Record 17 (1) 41-46.

- [15] Dokas P., Ertöz L., Vipin Kumar., Srivastava J., Tan P., "Data Mining for Network Intrusion Detection". National Science Foundation Workshop on Next Generation Data Mining, USA 2002
- [16] Xi Z., Sun J., Wenjie L., "Intrusion Detection using Fuzzy Window Markov Model". CCECE Niagra Falls, Canada 2004
- [17] Gomez J., Dasgupta D., "Evolving Fuzzy classifiers for Intrusion Detection". Proceedings of 2002 IEEE Workshop in Information Assurance, USA NY 2002
- [18] Manic M., and Wilamowski, "Fuzzy preference approach for Computer Network Attack Detection". IEEE conference on Fuzzy systems 2001
- [19] Yao TJ, Zhao, Saxton A study on fuzzy intrusion detection, Unpublished work, University of Regina, Canada, 2003
- [20] Y. Chen and J.Z. Wang, "Support Vector Learning for Fuzzy Rule-Based Classification Systems", IEEE Transaction on Fuzzy Systemss, 2003
- [21] Dickerson J E., Dickerson J A., "Fuzzy Network Profiling for Intrusion Detection" Proceedings of NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society, Atlanta 2000.
- [22] SNORT www.snort.org
- [23] SNORT INLINE <http://snort-inline.sourceforge.net/>
- [24] Zadeh, L. A. "The concept of a linguistic variable and its application to approximate reasoning, Parts 1, 2, and 3," Information Sciences, 1975.
- [25] SANS Global incident Analysis center Egress Filtering <http://www.sans.org/y2k/egress.htm>