ISBN 978-93-5156-328-0
International Conference of Advance Research and Innovation (ICARI-2014)

Advanced Intrusion Detection System Using Data Mining

Harshal R. Borse *, Abhijeet G. Garud, Jagruti S. Chopada
Department of Computer Engineering, University of Pune, Sandip Institute of Engineering & Management, Nashik, Maharashtra, India

Article Info

Article history:
Received 2 January 2014
Received in revised form
10 January 2014
Accepted 20 January 2014
Available online 1 February 2014

Keywords

DDoS Attacks
Data Mining,
IDS
Data Set
Apriori Algorithms
Internet Episode Rule
Pruning Techniques
Selective Inventory Control
Inventory Reduction

Abstract

The number of hacking and intrusion incidents is increasing alarmingly each year as new technology rolls out. In this paper report, we designed Intrusion Detection System (IDS) that implements predefined algorithms for identifying the attacks over a network. In this paper we discuss the term Intrusion Detection System using Data Mining which is generally used with the net working applications where the hacking attempts are made by the hackers. The key ideas are to use data mining techniques to discover consistent and useful patterns of system features that describe program and user behavior and use the set of relevant system features to compute (inductively learned) classifier that can recognize anomalies and known intrusions.

1. Introduction

One of the main challenges in the security management of large-scale high speed networks is the detection of suspicious anomalies in network traffic patterns due to Distributed Denial of Service (DDoS) attacks or worm propagation. A secure network must provide the following:

- Data confidentiality: Data that are being transferred through the network should be accessible only to those that have been properly authorized.
- Data integrity: Data should maintain their integrity from the moment they are transmitted to the moment they are actually received. No corruption or data loss is accepted either from random events or malicious activity.
- Data availability: The network should be resilient to Denial of Service attacks. [1]

1.1 What is IDS?

Intrusion detection is often used as another wall to protect computer systems. Intrusion detection (ID) is defined as the problem of identifying individuals Who are using a computer system without authorization (i.e., `crackers') and those who have legitimate access to the system but are abusing their privileges. The goal of IDS is to detect malicious traffic. In order out going traffic. There are several approaches on the implementation of IDS. Among those, two are the most popular.

Corresponding Author
E-mail address:bharshalborse4@gmail.com

1.2 Anomaly detection

This technique is based on the detection of traffic anomalies. The deviation of the monitored traffic from the normal profile is measured. Various different implementations of this technique have been proposed, based on the metrics used for measuring traffic profile deviation.

1.3 Misuse/Signature detection

This technique looks for patterns and signatures of already known attacks in the network traffic. A constantly updated database is usually used to store the signatures of known attacks. The way this technique deals with intrusion detection resembles the way that antivirus software operates. Data mining can help improve intrusion detection by addressing each and every one of the above mentioned problems.

1.4 Data mining

Data mining (DM), also called Knowledge-Discovery and Data mining is, at its core, pattern finding. Data miners are experts at using specialized software to find regularities (and irregularities) in large data sets.

2. Literature Survey

Computer Forensics, which views computer systems as scenes of a crime, is computer security technologies that analyze what attackers have done. Most of their applications focus on how to identify malicious network behaviors and the characteristics of attack packets, and the way to identify attack patterns based on their analysis.

Abdullah et al. [5] used package dump tools, such as tcpdump and pcap, to collect and analyze network packets and to identify network attacks from different network states and packets distribution.

Yu et al. [6] provided another example of integrating computer forensics with IDS. A knowledge-based system was deployed to collect forensic features frommalicious network behaviors. This system performed excellently in improving the hit rate of intrusion alerts.

3. System Architecture

In this section, we first introduce the data mining concept used for advanced intrusion and anomaly detection. Then we described the AIDS architecture, the ADS design, and the connection features used in ADS and automated signature generation and then prevention phase is discussed to avoid further similar attacks in future.

3.1 Detection Phase

In the proposed system architecture, the first step is to filter out the known attacks traffic by SNORT through signature matching either the database. The remaining traffic containing unknown or burst attacks fed to episode mining engine to generate frequent episode rule with different levels of support threshold.

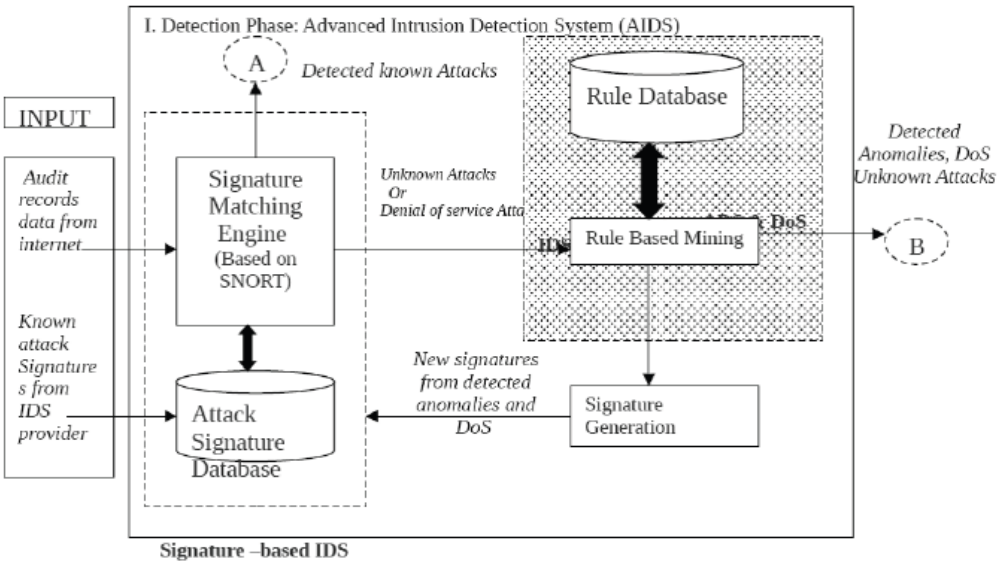

Fig: 1. Detection phase of AIDS

The anomalous episodes are used to generate signature which capture the anomalous behavior using a weighted frequent item set mining scheme. These signatures are then added to the SNORT database for future detection of similar attacks. Unknown, burst or multi-connection attacks are detected by ADS. The signature generation unit bridges two detection subsystems. This unite characterizes the detected anomalies and extracts their signatures. We build ADS by using the FER mining mechanism is described in next section. The new AIDS detects many novel attacks hidden in common Internet services, such as telnet, http, FTP, SMTP, e-mail, authentication, and so forth. The AIDS deployment appeals particularly to protect network based clusters of computers, resources inside internal networks (intranets), and computational grids.

3.2 Prevention phase

This is the last and most important phase of this project. In this phase, attacks detected by detection phase as shown in Fig 3.1 are given as input to this phase and after their correction, to suitable prevention policy can be applied to avoid the intrusion on the system. Fig 2 shows the prevention phase of this project and further selection gives the details about the functionality of each prevention phase policy and their usage choice for network administrator.

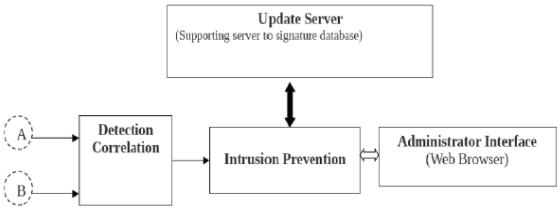

Fig: 2. Prevention phase of AIDS

Prevention policy selection used by network administrator during the prevention phase:

- Drop packets: The system allows the IDS to work in inline mode, it is able to drop or block a single packet, single session, or traffic now between the attack source and destination in real time, thwarting an attack in progress without affecting any other traffic.
- Terminate Session: The prevention phase allows for the initiation of TCP resets to targeted systems, attackers or both. The network security engineer can configure reset packets to be sent to the source and/or destination IP address.
- Modify Firewall policies: This allows user to reconfigure network firewall as an attacks occurs by temporarily challenging the user specified access control policy while alerting
  the security manager.
- Generate Alerts: This option enables an alert filter that allows network security engineer to sift out alerts based on the source or destination of the security events.

Log packets: Systems based on this architecture capture and log packets prior, during, or subsequent to the attack and redirect traffic to a spare system port for detailed forensic analysis. This packet information acts as a record of the actual ow of traffic that triggered the attack. When the data is viewed it is converted in to libpcap format for presentation. Tools like eternal, a network protocol analyzer for UNIX and windows, can be use to examine the packet log data for more detailed analysis of the detected event. The system architecture's response provides the basis for the product platform that security managers need to develop a system of actions alerts and logs its record. As mentioned earlier by examining this log data; it is possible to find out intrusive actions of the user and to reset their access rights for the network access.

4. Methodology for Development

In this section, complete details of the project system development are given. First part discusses the solution to reduce the unwanted episode rules generation by Appling pruning technologies on it. After that it focuses on the new proposed mining algorithm to eliminate uninteresting patterns discovery associated with frequent services at the time of low threshold applications. Also it give details on weighted apriori algorithm for generating signature forms anomalies detected. In the last session actual system implementation details are specified.

4.1 Internet episode rule and pruning techniques

An internet episode is represented by sequence of connections. An episode can be generated by sequence of connection events, such as, TCP, UDP, ICMP or other connections. An episode can be generated by legitimate user or malicious attackers. Frequently episodes are most resulted from normal user. A rare episode is likely caused by intruders. Our purpose is to build an AIDS that can distinguish the rare or abnormal episode from the normal or frequent episodes automatically. Following steps are used to differentiate such abnormal episode from the normal or frequent episodes automatically.

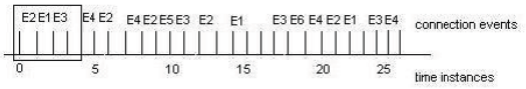

Fig: 3. Generation of frequent episode rule by scanning Stream of internet

Generation of Internet Episode Rule: The typical stream of internet traffic, represented by sequence of connections events label as E1, E2, E3 and so forth. These connection events are related to various internet service commands such as FTP, http, SMTP, authentication and so forth. Note that some events may repeat to appear in the sequence. The time instance of these connections, in seconds, is marked below the events. A frequent episode is a set of connection events exceeding the occurrence threshold in a scanning window. A FER is generated out of a

collection of frequent episodes. The FER is defined over episode sequences corresponding to multiple connection events in a roll. Base support traffic data mining: Most mining technique excludes infrequent traffic patterns. This will make the IDS ineffective in detecting rare network events. If we lower the support threshold, then a large number of uninteresting patterns associated with frequent services will be discovered. We introduce a new base support mining process to handle this problem. The process is specified in following algorithm. Our method is improved form the level wise algorithm by lee et al.. This is known as base support traffic data mining algorithm.

4.2 Weighted apriori algorithm for generating signature from detected anomalies

1. $\sum_{t \epsilon T} w_t$
2. k = i;
3. l_1 = { i|i ϵ I ^ wsup }; {find all weighted frequent 1 items }
4. Repeat
5. K= k+1;
6. C_k = aprior_gen (L_k_1); {Generate candidate item sets}
7. For each connection t ϵ T, do
8. C_t = subset (C_k,t); {candidates contained in t}
9. For each candidate item set c ϵ C_t, do
10. C.weight += w_t; {Add connection weight}
11. End for
12. End for
13. L_k={cϵ C_k| c.weight/w>=min_wsup};
14. Until L_k = NULL;
15. Return x =Ụ L_k.

5. Conclusion

This paper has presented a way of the data mining technique that has been proposed towards the enhancement of IDS. We have shown the way in which data mining has been known to aid the process of Intrusion Detection and the way in which the technique have been applied and evaluated by researcher. Finally, we pro-posed a data mining approach that we feel can contribute significantly in the attempt to create better and more effective Intrusion Detection Systems.

References

[1] A. A. Rao, P. Srinivas, B. Chakravarthy, K. Marx, P. Kiran, “A Java Based Network Intrusion Detection System (IDS) Session ENG in Proceedings of The 2006”, IJME INTERTECH Conference (2006) ,pp. 206-118

[2] S. P. Parekh, B. S. Madan, R. M. Tugnayat, “Approach For Intrusion Detection System Using Data Mining” Journal of Data Mining and Knowledge Discovery 2012, pp.-83-87.

[3] M. Thottan, C. Y. Ji, “Anomaly Detection in IP networks”, IEEE Transactions on Signal Processing, vol.51, no.8, Aug. 2003, pp.2191-2204

[4] G. Carl, G. Kesidis, R. R. Brooks, S. Rai, “Denial-of-Service Attack-Detection Techniques”, the IEEE Internet Computing, Jan. /Feb. 2006, pp. 82-89

[5] K. Abdullah, C. Lee, G. Conti, J. A. Copeland, “The IEEE Workshop on Information Assurance Workshop”, 2005

[6] J. Q. Yu, Y. V. R. Reddy, S. Selliah, S. Kankanahalli, S. Reddy, V. Bharadwaj, “TRINETR: An Intrusion Detection Alert Management System”, 235-240, 2004