Volume : 2 | Issue : 8 | August 2013 • ISSN No 2277 - 8179

Implementation of Layered CRF's For Intrusion Detection

Engineering

KEYWORDS :

K. SRI LAKSHMI M.Tech Student (SE), Sri Mittapalli College of Engineering

V. KESAVA KUMAR Associate Professor, Dept of CSE, Sri Mittapalli College of Engineering

ABSTRACT

Intrusion detection faces a number of challenges; an intrusion detection system must reliably detect malicious activities in a network and must perform efficiently to cope with the large amount of network traffic. In this paper, we address these two issues of Accuracy and Efficiency using Conditional Random Fields and Layered Approach. We demonstrate that high attack detection accuracy can be achieved by using Conditional Random Fields and high efficiency by implementing the Layered Approach. Intrusion detection is one of the high priority and challenging tasks for network administrators and security professionals. More sophisticated security tools mean that the attackers come up with newer and more advanced penetration methods to defeat the installed security systems. Finally, our system has the advantage that the number of layers can be increased or decreased depending upon the environment in which the system is deployed, giving flexibility to the network administrators. The areas for future research include the use of our method for extracting features that can aid in the development of signatures for signature-based systems. The signature-based systems can be deployed at the periphery of a network to filter out attacks that are frequent and previously known, leaving the detection of new unknown attacks for anomaly and hybrid systems.

I. INTRODUCTION

An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system.

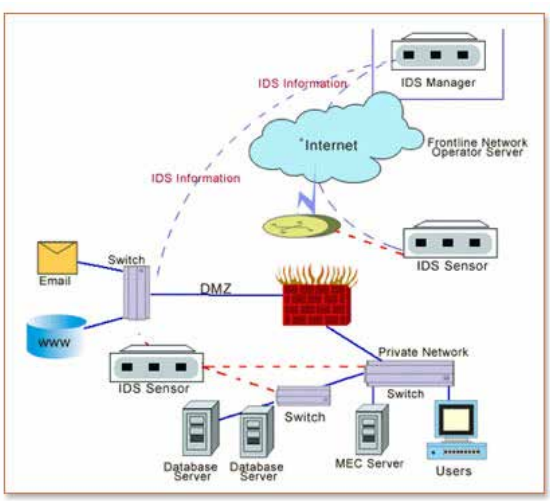

Fig 1: Intrusion Detection

Intrusion detection functions include:
Monitoring and analyzing both user and system activities.
Analyzing system configurations and vulnerabilities.
Assessing system and file integrity.
Ability to recognize patterns typical of attacks.
Analysis of abnormal activity patterns.
Tracking user policy violations.

ID systems are being developed in response to the increasing number of attacks on major sites and networks, including those of the Pentagon, the White House, NATO, and the U.S. Defence Department. The safeguarding of security is becoming increasingly difficult, because the possible technologies of attack are becoming ever more sophisticated; at the same time, less technical ability is required for the novice attacker, because proven past methods are easily accessed through the Web.In a WSN, there are two ways to detect an object (i.e., an intruder): single-sensing detection and multiple-sensing detection. In the single-sensing detection, the intruder can be successfully detected by a single sensor. On the contrary, in the multiple-sensing detection, the intruder can only be detected by multiple collaborating sensors .In some applications, the sensed information provided by a single sensor might be inadequate for recognizing the intruder. It is because individual sensors can only sense a portion of the intruder. For example, the location of an intruder can only be determined from at least three sensors' sensing. In view of this, we analyze the intrusion detection problem under two application scenarios: single-sensing detection and multiple-sensing detection.

II. PREVIOUS WORK

The most closely related work, to our work, is of Lee et al.[1], [2], and [3]. They, however, consider a data mining approach for mining association rules and finding frequent episodes in order to calculate the support and confidence of the rules separately. Instead, in our work, we define features from the observations as well as from the observations and the previous labels and perform sequence labeling via the CRFs to label every feature in the observation. This setting is sufficient for modeling the correlation between different features of an observation. We also compare our work with [4], which describes the use of maximum entropy principle for detecting anomalies in the network traffic. The key difference between [6] and our work is that the authors in [3] use only the normal data during training and build a baseline system, i.e., a behaviour-based system, while we train our system with both the normal and the anomalous data, i.e., we build a hybrid system. Second, the system in [4] fails to model long-range dependencies in the observations, which can be easily represented in our model. We also integrate the Layered Approach with the CRFs to gain the benefits of computational efficiency and high accuracy of detection in a single system. We compare the Layered Approach with the works in [8], [5], and [4]. The authors in [8] describe the combination of "strong" classifiers using stacking, where the decision tress, naive Bayes, and a number of other classification methods are used as base classifiers. The authors show that the output from these classifiers can be combined to generate a better classifier rather than selecting the best one. In [7], the authors use a combination of "weak" classifiers. The individual classification power of weak classifiers is slightly better than random guessing. The authors show that a number of such classifiers when combined using simple majority voting mechanism, provide good classification. In [10], the authors apply a combination of anomaly and misuse detectors for better qualification of analyzed events. However, our work is not based upon classifier combination. Combination of classifiers is expensive with regard to the processing time and decision making. The purpose of classifier combination is to improve accuracy. Rather, our system is based upon serial layering of multiple hybrid detectors.

III. MODULE ANALYSIS

Constructing Network Security:
In this module, we are going to connect the network each node

is connected the neighboring node and it is independently deployed in network area. And also deploy the each port no is authorized in a node. Intrusion detection as defined by the Sys Admin, Audit, Networking, and Security (SANS) Institute is the art of detecting inappropriate, inaccurate, or anomalous activity. Today, intrusion detection is one of the high priority and challenging tasks for network administrators and security professionals.

Randomized Field Detection:
In this module, browse and select the source file. And selected data is converted into fixed size of packets. And the packet is send from source to detector. Conditional models are probabilistic systems that are used to model the conditional distribution over a set of random variables. Such models have been extensively used in the natural language processing tasks. Conditional models offer a better framework as they do not make any unwarranted assumptions on the observations and can be used to model rich overlapping features among the visible observations.

Layered Approach for Intrusion Detection:
We now describe the Layer-based Intrusion Detection System (LIDS) in detail. The LIDS draws its motivation from what we call as the Airport Security model, where a number of security checks are performed one after the other in a sequence. Similar to this model, the LIDS represents a sequential Layered Approach and is based on ensuring availability, confidentiality, and integrity of data and (or) services over a network. The goal of using a layered model is to reduce computation and the overall time required to detect anomalous events. The time required to detect an intrusive event is significant and can be reduced by eliminating the communication overhead among different layers. This can be achieved by making the layers autonomous and self-sufficient to block an attack without the need of a central decision-maker. Each layer in the LIDS framework is trained separately and then deployed sequentially. We define four layers that correspond to the four attack groups mentioned in the data set.

Find Authorize & Unauthorized Port:
The intrusion detection is defined as a mechanism for a WSN to detect the existence of inappropriate, incorrect, or anomalous moving attackers. In this module check whether the path is authorized or unauthorized. If path is authorized the packet is send to valid destination. Otherwise the packet will be deleted. According port no only we are going to find the path is authorized or Unauthorized.

Constructing Inter-Domain Packet Filters:
If the packet is received from other than the port no it will be filtered and discarded. This filter only removes the unauthorized packets and authorized packets send to destination.

Receiving the Valid Packet:
In this module, after filtering the invalid packets all the valid Packets will reach the destination.

IV. SYSTEM IMPLEMENTATION
Layered approach for intrusion detection:
The Layer-based Intrusion Detection System (LIDS) in detail. The LIDS draws its motivation from what we call as the Airport Security model, where a number of security checks are performed one after the other in a sequence. Similar to this model, the LIDS represents a sequential Layered Approach and is based on ensuring availability, confidentiality, and integrity of data and (or) services over a network. Fig. 3 gives a generic representation of the framework. In order to make the layers independent, some features may be present in more than one layer. The layers essentially act as filters that block any anomalous connection, thereby eliminating the need of further processing at subsequent layers enabling quick response to intrusion. The effect of such a sequence of layers is that the anomalous events are identified and blocked as soon as they are detected

The discussed two main requirements for an intrusion detection system; accuracy of detection and efficiency in operation. We select features for each layer based upon the type of attacks that the layer is trained to detect as below.

1. Probe layer:
The probe attacks are aimed at acquiring information about the target network from a source that is often external to the network. Hence, basic connection level features such as the "duration of connection" and "source bytes" are significant while features like "number of files creations" and "number of files accessed" are not expected to provide information for detecting probes.

2**. DOS layer:**
For the DOS layer, traffic features such as the "percentage of connections having same destination host and same service" and packet level features such as the "source bytes" and "percentage of packets with errors" are significant. To detect DOS attacks, it may not be important to know whether a user is "logged in or not."

3. R2L layer:
The R2L attacks are one of the most difficult to detect as they involve the network level and the host level features. We therefore selected both the network level features such as the "duration of connection" and "service requested" and the host level features such as the "number of failed login attempts" among others for detecting R2L attack.

4. U2R layer (User to Root attacks):
The U2R attacks involve the semantic details that are very difficult to capture at an early stage. Such attacks are often content based and target an application. Hence, for U2R attacks, selected features such as "number of file creations" and "number of shell prompts invoked," while we ignored features such as "protocol" and "source bytes."

Algorithm
Training
Step 1: Select the number of layers, n, for the complete system.

Step 2: Separately perform features selection for each layer.

Step 3: Train a separate model with CRFs for each layer using the features selected from Step 2.

Step 4: Plug in the trained models sequentially such that only the connections labeled as normal are passed to the next layer.

Testing
Step 5: For each (next) test instance perform Steps 6 through 9.

Step 6: Test the instance and label it either as attack or normal.

Step 7: If the instance is labeled as attack, block it and identify it as an attack represented by the layer name at which it is detected and go to Step 5. Else pass the sequence to the next layer.

Step 8: If the current layer is not the last layer in the system, test the instance and go to Step 7. Else go to Step 9.

Step 9: Test the instance and label it either as normal or as an attack. If the instance is labeled as an attack, block it and identify it as an attack corresponding to the layer name.

CONCLUSION
Based on the incongruity intrusion detection principle our system is compared with other well known methods. By considering normal mathematical method, data mining and machine learning approaches anomaly-based systems largely detect deviations from the learnt normal data. How ever decision trees and naive Bayes are well verse for their performance but the results of our approach shows that Layered CRFs performs better than those methods and the cause is CRFs do not consider the observation features to be independent. For our methods mathematical testing also provided higher confidence in detec-

tion accuracy. Thus we proved that our system is strong and can handle noisy free data providing with high performance.

REFERENCE

[1] W. Lee and S. Stolfo, "Data Mining Approaches for Intrusion Detection," Proc. Seventh USENIX Security Symp. (Security '98), pp. 79-94, 1998. | [2] W. Lee, S. Stolfo, and K. Mok, "Mining Audit Data to Build Intrusion Detection Models," Proc. Fourth Int'l Conf. Knowledge Discovery and Data Mining (KDD '98), pp. 66-72, 1998. | [3] W. Lee, S. Stolfo, and K. Mok, "A Data Mining Framework for Building Intrusion Detection Model," Proc. IEEE Symp. Security and Privacy (SP '99), pp. 120-132, 1999. | [4] M. Sabhnani and G. Serpen, "Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context," Proc. Int'l Conf. Machine Learning, Models, Technologies and Applications (MLMTA '03), pp. 209-215, 2003. | [5] H. Shah, J. Undercoffer, and A. Joshi, "Fuzzy Clustering for Intrusion Detection," Proc. 12th IEEE Int'l Conf. Fuzzy Systems (FUZZ-IEEE '03), vol. 2, pp. 1274-1278, 2003. | [6] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning," Introduction to Statistical Relational Learning, 2006. | [7] R. Bace and P. Mell, Intrusion Detection Systems, Computer Security Division, Information Technology Laboratory, Nat'l Inst. of Standards and Technology, 2001. | [8] T.G. Dietterich, "Machine Learning for Sequential Data: A Review," Proc. Joint IAPR Int'l Workshop Structural, Syntactic, and Statistical Pattern Recognition (SSPR/SPR '02), LNCS 2396, pp. 15-30, 2002. | [9] P. Dokas, L. Ertoz, A. Lazarevic, J. Srivastava, and P.-N. Tan, "Data Mining for Network Intrusion Detection," Proc. NSF Workshop Next Generation Data Mining (NGDM '02), pp. 21-30, 2002. | [10] Y. Du, H. Wang, and Y. Pang, "A Hidden Markov Models-Based Anomaly Intrusion Detection Method," Proc. Fifth World Congress on Intelligent Control and Automation (WCICA '04), vol. 5, pp. 4348-4351, 2004. |