International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-3 Issue-5 September-2012

A Study Paper on IDS Attack Classification Using Various Data Mining Techniques

Sneha Kumari[1], Maneesh Shrivastava[2]
Department of Information Technology, LNCT, Bhopal, India[1,2]

Abstract

Over the past several years, the Internet environment has become more complex and untrusted. Enterprise networked systems are inevitably exposed to the increasing threats posed by hackers as well as malicious users internal to a network. IDS technology is one of the important tools used now-a-days, to counter such threats. Various IDS techniques has been proposed, which identifies and alarms for such threats or attacks. Data mining provides a wide range of techniques to classify these attacks. The paper provides a comparative study on the attack detection rate of these existing classification techniques.

Keywords

Intrusion Detection system, Attack, Data Mining, Classification, Algorithm

1. Intrusion Detection System

Intrusion detection is the process of trying to find out activities that violate security policy when they are taking place in computer networks and systems [1]. Since its invention, intrusion detection has been one of the key elements in achieving information security. It acts as the second-line defence which supplements the access controls. When the controls failed, the intrusion detection systems should be able to detect it real-time and warn the security officers to take prompt and appropriate actions.

The figure1 given below shows the generic architecture of IDS [2]. The event generator is the source of audit trail data and it is responsible for collecting data for analysis. Audit trail data can be network traffic logs, system call logs, user activities history, etc. The data collection policy indicates which kinds of data to get and rules to pre-process the data. The analysing unit implements the detection algorithms and looks for suspicious activities from the audit data.

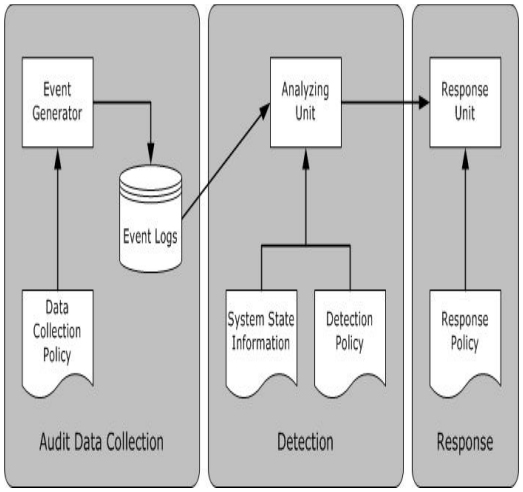

Figure 1: Components in generic IDS

The detection policy in Figure 1 describes how intrusions are detected. This is where the attack signatures and operating parameters are stored. The outcome of analysing unit is alarms of suspicious activities. Current state information (e.g. partially matched suspicious activities and current system status) is stored and is provided to the analysing unit during detection process as needed. The response unit will react to the alarms generated from the analysing unit. The responses can be manual or automated actions.

Where to do Intrusion Detection

On the Border or switched network port

- Distributed
- Host Based

Challenges to Intrusion Detection

There are many challenges to doing intrusion detection:

- False positives: Administrators can be totally bogged down by false positives which are essentially warnings about things.
- Learning curves: Intrusion detection can be a technically challenging environment that may require a substantial learning curve.

- Large Logs: Logs of events are useless unless that are looked at via some mechanism.
- Placement of IDS: Where do you place your IDS in order to effectively catch intrusion attempts [3]?

In Summary you need to run IDS to:
Prevent Hackers/Crackers from abusing your systems
Help Prevent viruses and Worms
Prevent Bandwidth theft
Actively act against threats
Keep you informed about the general state of the network

Shortfalls with current IDS:-

1. Variants
2. False positives
3. False negatives
4. Data overload [7].

2. Types of Attacks

Depending upon the harm and measure of threats attacks have been categorized into following categories:

A. Probing: It is a class of attack where an attacker scans a network to gather information or find known vulnerabilities. An attacker with a map of machines and services that are available on a network can use the information to look for exploits. There are different types of probes: some of them abuse the computer's legitimate features; some of them use social engineering techniques. This class of attacks is the most commonly heard and requires very little technical expertise.

B. Denial of service attacks: DoS is a class of attacks where an attacker makes some computing or memory resource too busy or too full to handle legitimate requests, thus denying legitimate users access to a machine There are different ways to launch DoS attacks: by Abusing the computers legitimate features; by targeting the implementations bugs; or by exploiting the system's misconfigurations. DoS attacks are classified based on the services that an attacker renders unavailable to legitimate users.

C. User to root attacks: User to root exploits are a class of attacks where an attacker starts out with access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system. Most common exploits in this class of attacks are regular buffer overflows, which are caused by regular programming mistakes and environment assumptions.

D. Remote to user attacks: A remote to user (R2L) attack is a class of attacks where an attacker sends packets to a machine over a network, then exploits machine's vulnerability to illegally gain local access as a user. There are different types of R2U attacks: the most common attack in this class is done using social engineering [4].

3. Introduction of Data Mining

Data mining is an ambiguous term that has been used to refer to the process of finding interesting information in large repositories of data. More precisely, the term refers to the application of special algorithms in a process built upon sound principles from numerous disciplines including statistics, artificial intelligence, machine learning, database science, and information retrieval [5].

Data mining algorithms are utilized in the process of pursuits variously called data mining, knowledge mining, data driven discovery, and deductive learning. Data mining techniques can be performed on a wide variety of data types including databases, text, spatial data, temporal data, images, and other complex data [6].

In summary data mining is:

- A hot buzzword for a class of techniques that find patterns in data
- A user-centric, interactive process which leverages analysis technologies and computing power
- A group of techniques that find relationships that have not previously been discovered
- Not reliant on an existing database
- A relatively easy task that requires knowledge of the business problem/subject matter expertise

4. How Can Data Mining Help In IDS

Data mining can help improve intrusion detection by adding a level of focus to anomaly detection. By identifying bounds for valid network activity, data mining will aid an analyst in his/her ability to

distinguish attack activity from common everyday traffic on the network.

A. Variants: Since anomaly detection is not based on pre-defined signatures the concern with variants in the code of an exploit are not as great since we are looking for abnormal activity versus a unique signature. An example might be a Remote Procedure Call (RPC) buffer overflow exploit whose code has been modified slightly to evade an IDS using signatures. With anomaly detection, the activity would be flagged since the destination machine has never seen an RPC connection attempt and the source IP was never seen connecting to the network.

B. False positives: In regards to false positives there has been some work to determine if data mining can be used to identify recurring sequences of alarms in order to help identify valid network activity which can be filtered out.

C. False negatives: Detecting attacks for which there are no known signatures. By attempting to establish patterns for normal activity and identifying that activity which lies outside identified bounds, attacks for which signatures have not been developed might be detected. An extremely simple example of how this would work would be to take a web server and develop a profile of the network activity seen to and from the system. Let us say the web server is locked down and only connections to ports 80 and 443 are ever seen to the server. Thus, whenever a connection to a port other than 80 or 443 is seen the IDS should identify that as an anomaly. While this example is quite simple this could be extended to profiling not only individual hosts, but entire networks, users, traffic based on days of the week or hours in a day, and the list goes on.

D. Data overload: The area where data mining is sure to play a vital role is in the area of data reduction. With current data mining algorithms there exists the capability to identify or extract data which is most relevant and provide analysts with different "views" of the data to aid in their analysis [7].

Data mining involves six common classes of tasks:

- **Anomaly detection:** The identification of unusual data records that might be interesting or data errors and require further investigation.
- **Association rule learning** (Dependency modeling)**:** Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.
- **Clustering :** is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
- **Classification:** is the task of generalizing known structure to apply to new data. For example, an email program might attempt to classify an email as legitimate or spam.
- **Regression:** Attempts to find a function which models the data with the least error.
- **Summarization:** providing a more compact representation of the data set, including visualization and report generation.

5. Various Classification Algorithms

Various classification algorithms have been proposed, which are:

- Linear classifiers
  - Fisher's linear discriminant
  - Logistic regression
  - Naive Bayes classifier
  - Perceptron
- Support vector machines
  - Least squares support vector machines
- Quadratic classifiers
- Kernel estimation
  - k-nearest neighbor
- Boosting
- Decision trees
  - Random forests
- Neural networks
- Bayesian networks
- Hidden Markov models
- Learning vector quantization

Some of the most commonly used classification algorithm which has been widely implemented in IDS is given below.

A. Decision trees: The well-known machine learning techniques. A decision tree is composed of three basic elements:

1. A decision node specifying a test attributes.
2. An edge or a branch corresponding to the one of the possible attribute values which means one of the test attribute outcomes.
3. A leaf which is also named an answer node contains the class to which the object belongs.

In decision trees, two major phases should be ensured:

1. **Building the tree:** Based on a given training set, a decision tree is built. It consists of selecting for each decision node the 'Appropriate' test attribute and also to define the class labelling each leaf.
2. **Classification:** In order to classify a new instance, we start by the root of the decision tree, then we test the attribute specified by this node. The result of this test allows moving down the tree branch relative to the attribute value of the given instance. This process will be repeated until a leaf is encountered. The instance is then being classified in the same class as the one characterizing the reached leaf. Decision trees have also been used for intrusion detection [4]. The decision trees select the best features for each decision node during the construction of the tree based on some well-defined criteria. One such criterion is to use the information gain ratio. [8]

B. Support Vector Machines: Support vector machines map real valued input feature vector to a higher dimensional feature space through nonlinear mapping and can provide real-time detection capability, deal with large dimensionality of data, and can be used for binary-class as well as multiclass classification SVM classify data by using these support vectors that outline the hyper plane in the feature space. This process will involve a quadratic programming problem, and this will get a global optimal solution. Suppose we have N training data points
{(x1, y1), (x2,y2), (x3, y3), ..., (xN , yN)},
Where x¡ € Rd and y¡ €{+1,−1}. Consider a hyper-plane defined by (w, b), where w is a weight vector and b is a bias. The classification of a new object x is done with N
f(x) =sign(w.x+ b) =sign($\Sigma\alpha_i y_i (x_i \cdot x) + b$)
The training vectors xi occurs only in the form of a dot product. For each training point, there is a Lagrangian multiplier α_i. The Lagrangian multiplier values αi reflects the importance of each data point. When the maximal margin hyper-plane is found, only points that lie closest to the hyper-plane will have $\alpha_i > 0$ and these points are called support vectors. All other points will have $\alpha_i = 0$. That means only those points that lie closest to the hyper plane, give the representation of the hypothesis/classifier. These data points serve as support vectors. Their values can be used to give an independent boundary with regard to the reliability of the hypothesis/classifier [9].

C. Bayesian network classifier: One of the most effective classifiers, in the sense that its predictive performance is competitive with state-of-the-art classifiers, is the so-called naive Bayesian classifier described. From training data the conditional probability of each attribute *Ai* given the class label *C*. Classification is then done by applying Bayes rule to compute the probability of *C* given the particular instance of *A*1*; : : : ; An*, and then predicting the class with the highest posterior Probability. This computation is rendered feasible by making a strong independence assumption: all the attributes *Ai* are conditionally independent given the value of the class *C*. By independence we mean probabilistic independence, that is, *A* is independent of *B* given *C* whenever Pr(*A*j*B;C*) = Pr(*A*j*C*) for all possible values of *A;B* and*C*, whenever Pr(*C*) > 0 [10].

Bayesian networks represent a new approach to detection and prevention of attacks in computer networks; the application of Bayesian networks in IDS solves the majority of problems present in previously discussed methods. Besides, Bayesian networks offer significant advantages which are not possible to implement using other methods. Relations between events are not given on the basis of expert knowledge but represent mutual relations between events in the domain under consideration. Network

training is not possible in a real environment; thus, systems for detection and prevention of attacks are not exposed to training by an intruder. Due to interrelationships of independent IDS, the previous knowledge on attacks in the entire network is synthesized and integrated into a unique system. Events used to estimate the probability of attacks are analysed at the network location where they happened; in this way, an unnecessary communication and processing overloading are avoided, and the problem of incompatible various control records (generated at different computer systems) does not exist. System using Bayesian network offers an unique advantage over other systems when one calculates the influence of newly produced events on the other observed events; accordingly, all data and rules used in other systems can be built into IDS based on Bayesian networks. Bayesian networks provide a full compatibility of corresponding software products without respect to platform used for execution; this fact can speed-up the development and application of standalone and distributed IDS [11].

D. Artificial Neural Network : An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

The ability of soft computing techniques for dealing with uncertain and partially true data makes them attractive to be applied in intrusion detection. Some studies have used soft computing techniques other than ANNs in intrusion detection. For example, genetic algorithms have been used along with decision trees to automatically generate rules for classifying network connections .An ANN is an information processing system that is inspired by the way biological nervous systems, such as the brain, process information. It is composed of a large number of highly interconnected processing elements (Neurons) working with each other to solve specific problems. Each processing element (neuron) is basically a summing element followed by an activation function. The output of each neuron (after applying the weight parameter associated with the connection) is fed as the input to all of the neurons in the next layer. The learning process is essentially an optimization process in which the parameters of the best set of connection coefficients (weighs) for solving a problem are found and includes the following basic steps :

1. Present the neural network with a number of inputs (vectors each representing a pattern)
2. Check how closely the actual output generated for a specific input matches the desired output.
3. Change the neural network parameters (weights) to better approximate the output [12].

Various above mentioned classification scheme are useful in IDS. The performance of these classifications can be detected by analysing the detection rate of these techniques. The table given below shows the comparison of attack detection rate in % of the above mentioned techniques [13, 4].

Table 1: Performance Evolution Of various classifications Technique

	NORMAL	PROBE	DOS	U2R	R2L
ANN	99.60	92.70	97.50	48	95.00
BNN	99.57	99.43	99.69	64	99.11
SVM	99.64	98.57	99.92	40	33.92
DT	99.64	99.86	96.83	68	84.19

6. Conclusion

This paper has presented a study of the various data mining techniques that have been proposed towards the enhancement of IDSs. We have shown the ways in which data mining has been known to aid the process of Intrusion Detection. Finally, in the last section, we proposed a comparative approach to get the performance rate of the four data mining approach. All the five attacks detection rate have been shown, on the basis of which we can say that

the u2r attacks detection rate is very less in all these classifier.

References

[1] Bace, R. "Intrusion Detection", Macmillan Technical Publishing. 2000.
[2] Lundin, E. and Jonsson, E. "Survey of research in the intrusion detection area, Technical Report 02-04", Department of Computer Engineering, Chalmers University of Technology, Göteborg, Sweden. January 2002.
[3] http://sdu.ictp.it/lowbandwidth/program/richard/Richard-intrusiondetection.pdf.
[4] Srinivas Mukkamalaa, Andrew H. Sunga and Ajith Abrahamb "Intrusion detection using an ensemble of intelligent paradigms"; www.elsevier.com/locate/jnca, January 2004.
[5] Jiawei Han and Micheline Kamber ,"Data mining: concepts and techniques", San Francisco: Morgan Kaufmann Publishers, 2001.
[6] Margaret Dunham, "Data Mining Introductory and Advanced Topics", ISBN: 0110888923, Prentice Hall, 2003.
[7] Manh Phung , "Intrusion Detection FAQ: Data Mining in Intrusion Detection", http://www.sans.org/security-resources/idfaq/data_mining.php, October 24, 2000.
[8] http://www.mendeley.com/research/naive-bayes-vs-decision-trees-in-intrusion-detection-systems/#page-1.
[9] Snehal A. Mulay, P.R. Devale and G.V. Garje, "Intrusion Detection System using Support Vector Machine and Decision Tree", International Journal of Computer Applications (0975 – 8887) Volume 3 – No.3, June 2010.
[10] Pedro Domingos and Michael Pazzani , "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss", Kluwer Academic Publishers. Manufactured in The Netherlands. , 111–133 (1997) °c 1997, Machine Learning, 1997.
[11] Milan Tuba, Dusan Bulatovic, Olga Mijkovic, Dana Simian, " Specific Attack Adjusted Bayesian Network for Intrusion Detection System", 9th WSEAS Int. Conf. on Mathematics & Computers In Biology & Chemistry (MCBC '08), Bucharest, Romania, June 24-26, 2008.
[12] Moradi, Mehdi, and Mohammad Zulkernine. "A neural network based system for intrusion detection and classification of attacks." In Proceedings of the 2004 IEEE international conference on advances in intelligent systems-theory and applications. 2004.
[13] Sandhya Peddabachigari, Ajith Abraham and Crina Grosan, Johnson Thomas, "Modeling intrusion detection system using hybrid intelligent systems", Journal of Network and Computer Applications 30 (2007) 114–112 in 2007.

Sneha Kumari, D/O Amar Nath Jha is from Deoghar, Jharkhand. Her DOB is 21/05/1987. She is pursuing M.tech degree in Information Technology from Rajiv Gandhi Proudyogiki Vishwavidyalaya (RGPV), Bhopal, India. Her research area is Ad Hoc Network and network Security.