Comprehensive study of classification model for intrusion detection system: Review

A. Ganesan[*1] and Dr. A. Kumar Kombaiya[2]
[1]PG & Research Department of Computer Applications, Hindusthan Arts College,
Coimbatore – 641 028, India.
[2]Department of Computer Science, Chikkanna Govt. Arts College, Thiruppur – 641 602, India.

Abstract: ***The Intrusions Detection is one of the essential components in network security. Intrusion detection system used for monitoring patterns of activities in user accounts and detects malicious activities. This paper presents an overview of classification model of intrusions detection system. There are many model and techniques have been used to implement intrusions detection system. These techniques are used as detector part of intrusions detection system (IDS) such as statistical model which includes the techniques of Mean, Standard deviations, multivarience, markoprocess model ,Data mining model includes classification ,association rule and clustering techniques and Artificial intelligence includes Fuzzy logic and knowledge expert system. To increase the performance of an Intrusions Detection System, IDS approaches may be used as combined together as more than one technique.***

Keywords: KDD, Intrusion, data mining, IDS

I. Introduction

Network and Internet technology has become a essential part in our day to day life. Business or personal Information transformation using network or internet is one of the major problems. Host based or Network based attacks are increasing frequently. Finally it makes huge financial loss to the organization security threats to the Information system have increased the essential of intrusion detection system.

The Intrusion detection system used to protect these networks from intruder (unauthorized person). In 1980 J.R Anderson presented the concepts that certain types of threats of the security could be identified a review of information contained in system's audit trail [4]. In 1985 denning and Neumann [3] has recommended the essential for effective intrusion detection mechanism.

The IDS (Intrusion detection system) is the process of intelligently monitoring the pattern of activity occurring in the network /internet and analyzing them for detecting an violation of the security policy. Parker [5] has defined six security issues such as confidentiality, integrity, Availability, utility, and possession of the network to be considered during the process of IDS. The intrusions are caused by attackers accessing the system from network/internet.

II. Intrusion detection model

The intrusion detection model can be grouped into three models. They are Statistical model, Data mining model and knowledge based expert system model. The statistical model contains statistical movements, such as mean, median and standard deviation. In data mining model, we can implement data mining algorithm such as association rule, decision tree, artificial neural network, clustering are implemented and third model is Knowledge based expert system model .

(a) Statistical Model:

The statistical model used to detect intrusion. This model use statistical properties such as mean, median, standard deviation and variance and test to determine whether observed user's Behaviour deviate from expected behaviour (from user profile).To build statistical model to find intrusion use statistical properties for user's normal activities for example – average login time of user. The statistical movement can provide accurate notification of abnormal activities. This method does not require prior knowledge of security flaws.

The activity profile is generated in statistical model by observing the system activities. An activity profile characterizes the behaviour of a given subject with respect to the given object.

From the training data set of the user behaviour the mean standard deviation can be computed
By the following formula.

Mean-vector [i] = 1/N $\sum_{j=1}^{N} Instance[i]$

Std-Deviation-Vector[i] = $\frac{1}{N-1} \sum_{j=1}^{N} (behavioural\ instance[i] - mean\ vector[i]^2)^{1/2}$

Where vector [i] is the i[th] behavioural instance of the vector.

The intrusion detection system assigning a score to the user normal behaviour. The IDS compute mean, standard deviation of the current abnormal user behaviour. If the current abnormal user behaviour value (ex: login time) greater than normal user behaviour value, it (IDS) will generate alarm.

The statistical anomaly based intrusion detection system establishes user behaviour profiles for normal activities and abnormal (current) activities. Then use these profiles are matched based on various methods to detect deviation from the normal behaviour. The statistical model can be
Classified as i) Operational model ii) markov process model iii) statistical movement model(mean, standard deviation) and iv) multivariate model.

In operational model an anomaly can be identified through the comparison of an observation with predefined limit. In this model the tolerant level of particular behaviour is known in advance.

The morkov process model use morkov chains and keep the track of an intrusion by examining. The system fixed intervals and maintain the record of state. If state changes takes place it computes the probability for that state at a given interval.

The multivariate model based on correlation between two or more variable. The intruder behaviour characterized with greater confidence by correlation (process time, login frequency).

b) Data Mining Model

Data mining is the process of extracting the data from large data store. The data mining Model used to provide better solution for intrusion by using 'pattern finding' method. The data Mining model deploy data mining algorithm such as association rule, clustering, classification, neural networks for intrusion detection.

Data mining model tends to reduce the amount data that must be retained for comparison of network activity. In data mining model the OLAP (online analytical process) engine provides an adequate interface for querying, summarises and aggregate information across different dimensions hierarchies. In data mining model, the term KDD knowledge discovery in data base is used to denote the process of extracting useful knowledge from large data sets, The KDD process uses data mining technique along any required pre and post processing to extract high level knowledge. The steps of KDD developing the application domain, data integration and selection, data mining, pattern evaluation and knowledge representation.

Data mining methods predict future trends and behaviours, it helps organizations to make proactive knowledge driven decisions. In this model we can use data mining (algorithm)method such as associates rules ,classifications ,clustering to extract the information needed to conclude intrusions are occurred or not.

Association Rule

The association rule was formulated by Agrawal in 1993 and is often referred to as the market basket problem. The task is to find relationship between the various items within these baskets. This association rule method used to detect the intrusion in network. An association rule is an expression of the form X => Y. where X and Y are the set of behaviour in user profile (example: login time, file usage, file type, location).The intuitive meaning of such rule is the event of the database which contain X tends to contain Y. The goal is to discover all the rules that have the **support** and **confidence** greater than or equal to the minimum support and confidence.

Let B = {b1, b2.....bn} be a set of behaviour in user profile. Let '**D'** the database (training data to be tested for intrusion) be a set of user profile with this behaviour. Where user '**U'** is a set of behaviour **U** supports a behaviour x, if x is in **U,** it is said to support a subset of behaviour x if **U** (users) support each behaviour x in X. X => Y has support in preset **D** is S% of events in **D** supports **X U Y.**

Support means how often X and Y occurs together as a percentage of the total events. Confidence measure how much a particular behaviour is depends on another. Thus the association rules used to find correlation between the behaviour (attributes).

Classification

Classification is one of the methods in data mining model. It involves finding rules that partition the data into disjoint groups. Using this method we can classify each instance (event) as normal or a particular kind of intrusion. The aim of classification to learn from class-labelled training instance for predicting class of new or previously unseen data. The new data is classified based on training set. The decision tree is the classification approach .this type of method is used to reduce the number of records to be examined individually .when huge amount of test records are involved .The testing phase of classification based technique is fast , since each test instance need to be compared against pre computed class.

Clustering

The clustering method is one of the data mining algorithms. Clustering is a method of grouping data so that the data in each group share similar trends and patterns. It attempts to automatically partition the data space into set of region or clusters.

The clustering is an unsupervised technique for finding pattern unlabeled data with many number of attributes. Using the clustering approach data intrusions are examined whether data instance having the same classification (type of attack/normal) or it is closes to each other under some reasonable metric. This method able to detect a particular subset of intrusion attack. The k- Means clustering method used to produce more accurate clustering, but it has greater time complexity.

c) Expert System Based Knowledge Based Model:

Intrusion can be detected by using another one method artificial intelligence based expert system model. This system act as human expertise. Expert systems are designed to solve complex problems by reasoning through bodies of knowledge, represented mainly as if-then-rules through conventional procedural code.

The components of experts system include knowledge base which contains possible user profile as logic representations. Inference engine is another component. It is used to derive the knowledge from knowledge base whenever the explanation is needed. The knowledge base will provide explanation. The Third component of expert system is user interface. It is used to access the knowledge base and to provide the query to the system. In knowledge base we can specify the user profile, time and date of access, user location, user id and type of access. The current time, location could be composed to the user profile to determine if the user is original user of the user id verified in the profile.

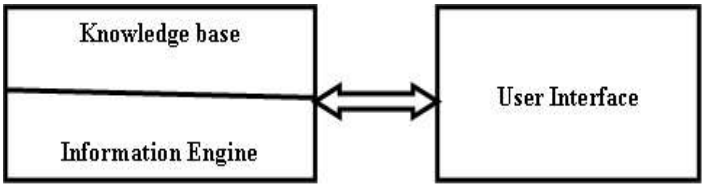

Figure 1: Expert system based model

The expert system is called as knowledge based intrusion detection system. The knowledge contains about specific attack and vulnerabilities the system. In knowledge base knowledge/fact is expressed as rule that describe suspicious behaviour based on knowledge of past intrusion, known system vulnerabilities. The rule describe suspicious behaviour that is independent of whether a user deviating from past behaviour pattern. The combination of the rule based component and the statistical based components gives greater detective power than either component by itself.

5. Conclusion

This paper has presented various model of intrusion detection like statistical model and its method, data mining model and Expert system based knowledge based model. In statistical oment (mean, standard deviation) based model does not require prior knowledge of security flaws. The reason using data mining model for intrusion detection system is the enormous volume of existing and newly appearing network data. That will be useful for future pattern generation.

In this paper we presented a detailed review of some important anomaly based intrusion detection model and their methods. The choice of model depends on various factors such as type of anomalies, data type, behaviour, working environment, security level and computational cost required.

5. References

[1] Andersion, J. (1995) An Introduction to neural networks, Cambridge: MIT Press.
[2] Canady J. (1998). Artificial neural networks for Misuse detection, Proceedings of the 1998 Nation information systems security Conference (NISSC'98), pp- 443-456, Arlington, VA.
[3] R. Agrawal, T Imielinski, and A. Ss. 1993. Mining Association Rules between Sets of Items in Large Databases. In Proceedings of the ACM SIGMOD Int'l Conf. on the Management of Data, pp. 207-216.
[4] K. Scarfone, P. Mell, Special Publication 800-94: Guide to Intrusion Detection and Prevention Systems (IDPS), National Institute of Standards and Technology (NIST) (2007).
[5] T. Abbes, A. Bouhoula, and M. Rusinowitch, Protocol analysis in intrusion detection using decision tree, Proceedings of International Conference on Information Technology: Coding and Computing (ITCC), vol. 1, 2004.
[6] A.A.E. Ahmed and I. Traore, Detecting computer intrusions using behavioral biometrics, Third Annual Conference on Privacy, Security and Trust (PST), 2005.
[7] Balajinath and SV Raghavan, Intrusion detection through learning behavior model, Computer Communications 24 (2001), no. 12, 1202–1212.
[8] S. Antonatos, K.G. Anagnostakis, and E.P. Markatos, Generating realistic workloads for network intrusion detection systems, ACM SIGSOFT Software Engineering Notes 29 (2004),No. 1, 207–215.
[9] O. Depren, M. Topallar, E. Anarim, and M.K. Ciliz, An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks, Expert systems with Applications 29 (2005), no. 4, 713–722.
[10] AK Ghosh, J. Wanken, and F. Charron, Detecting anomalous and unknown intrusions against programs, Proceedings of the 14th Annual Computer Security Applications Conference (ACSAC' 98), 1998, pp. 259–267.
[11] Ko, D. A. Frincke, T. Goan, L. T. Heberlein, K. Levitt, B. Mukherjee, C. Wee, "Analysis of an Algorithm for Distributed Recognition and Accountability", 1st ACM Conference on Computer and Communication Security, pp. 154-164, 1993.
[12] L. Gasser, "An overview of DAI", Kluwer Academic Publisher, Boston 1992.
[13] M. Crosbie, E. H. Spafford, "Applying Genetic Programming to Intrusion Detection", Technical report, Computer Sciences Department, Purdue.
[14] Balasubramaniyan, J. O. Garcia-Fernadez, D. Isacoff, E. H. Spafford, D. Zamboni, "An Architecture for Intrusion Detection using Autonomous Agents", Technical report Coast-TR-98-05, Computer Sciences Department, Purdue University.
[15] E. Denning, "An Intrusion-Detection Models", IEEE Transactions on Software Engineering, Volume SE-13, No. 2, February 1987.