© 2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Detection of Intrusion Using Decision Tree Based Data Mining Technique

M. Chandrasekhar Varma[1], B Srinivasa S P Kumar[2]

[1]Assistant professor, Department of CSE, D.N.R College of Engineering and technology, Bhimavaram, Andhra Pradesh, India

[2]Assistant Professor, Department of MCA, CBIT, Hyderabad, Telangana, India

ABSTRACT

Now a days computer attack has turned out to be exceptionally normal. Despite the fact that there are numerous current instruments for Intrusion detection, yet the primary issues is the security and precision of the system. This paper is presented with J48 decision tree data mining techniques as an intrusion detection system and ORNL (Oak Ridge National Laboratories) data set. This paper demonstrates that Decision trees gives higher precision than other regulated strategy.

Keywords: Decision tree, Data Mining, J48 algorithm, Intrusion Detection.

I. INTRODUCTION

These days, numerous associations and organizations utilize Internet benefits as their correspondence and commercial center to work together, for example, at EBay and Amazon.com site. Together with the development of computer arrange exercises, the developing rate of system attacks has been progressing, affecting to the accessibility, classification, and uprightness of basic data. In this manner a system must utilize at least one security devices, for example, firewall, antivirus, IDS and Honey Pot to keep vital data from criminal endeavors. Because of expanded number of web clients there is an issue because of intrusion which may harm data and data put away in computer server or data base server. So we require a channel which can channel vindictive data and ordinary data.

Intrusion detection is the way toward checking and breaking down the occasions happening in a computer system with a specific end goal to recognize indications of security issues. The intrusion detection and other security advancements, for example, cryptography, validation and firewalls has picked up in significance in most recent couple of years [3].

As system based computer systems assume progressively indispensable parts in current society, they have moved toward becoming intrusion detection systems give following three fundamental security capacities Data classification, Data respectability, Data accessibility [12].

Classification (or mystery) implies that data is revealed just as per strategy, respectability implies that data isn't pulverized or adulterated and that the system performs effectively, accessibility implies that system administrations are accessible when they are required [13].

There are two sorts of intrusion detection techniques: Misuse and Anomaly. Abuse locators investigate system action, searching for occasions or sets of occasions that match a predefined example of occasions that depict a known attack. As the examples comparing to known attacks are called marks, abuse detection is some of the time called "signature-based

detection." Anomaly identifiers recognize anomalous unordinary conduct (peculiarities) on a host or system. They work on the supposition that attacks are unique in relation to "typical" (real) movement and can in this manner be distinguished by systems that recognize these distinctions [4]. Peculiarity detection system screens the conduct of a system and banner noteworthy deviations from the typical action as an inconsistency. Inconsistency detection is utilized for distinguishing attacks in a computer systems, vindictive exercises in a computer systems, abuses in a Web-based systems [14].

In this paper we have utilized data mining way to deal with intrusion detection. This paper for the most part centers around the mark based intrusion detection systems and presents an approach to recognize examples of hurtful attacks via preparing the system on a database and testing the same. So as to help the preparation and testing the ORNL dataset is utilized, which comprises of various sorts of system associations named with the class. A model with high exactness will be attempted to create .Model will be prepared and tried on the typical and known attacks. Whatever remains of the papers comprise following areas as took after. Segment 2presents an audit of related work. Segment 3 bargains our proposed work. Area 4 presents the fundamental idea of strategy we utilized .Section 5 depicts result. And last segment closes the paper.

II. LITERATURE SURVEY

Intrusion detection began in 1980's and from that point forward various techniques have been acquainted with assembled intrusion detection systems [2]. Right now constructing a compelling ID is a gigantic learning building errand. System manufacturers transfer on their instinct and experience to choose the factual measures for inconsistency detection. Specialists initially dissect and order attack situations and system vulnerabilities, and hand-code the comparing guidelines and examples for abuse detection. On account of the manual and Adhoc nature of the improvement procedure, current IDSs have constrained extensibility and flexibility. Numerous IDSs just handle one specific review data source, and their updates are costly and moderate [5][6].

Authors[7] proposed a multi-Layer intrusion detection. There trial comes about demonstrated that the proposed multi-layer show utilizing C5 decision tree accomplishes higher grouping rate precision, utilizing highlight choice by Gain Ratio, and less false alert rate than MLP and gullible Bayes. Utilizing Gain Ratio improves the precision of U2R and R2L for the three machine learning techniques (C5, MLP and Naïve Bayes) fundamentally. MLP has high characterization rate when utilizing the entire 41 includes in Dos and Probe layers.

Authors[5] This paper demonstrates the usage by review intrusion detection as a data mining issue. A standout amongst the most widely recognized data mining approaches i.e. arrangement through decision trees has been embraced to recognize intrusion detection designs. There is a restriction that it can't distinguish obscure attacks.

Authors[8] assessed the impact of quality pre-choice utilizing Statistical techniques on certifiable kddcup99 data set. Test result demonstrates that precision of the C4.5 classifier could be enhanced with the strong pre-choice approach when contrast with conventional element choice techniques yet the main confinement in this exploration paper is actualizing right characteristic determination measure in C4.5 decision tree algorithm.

Authors[9] This paper closes numerous grouping techniques that were beforehand proposed to take care of the natural IDS issues. Where, the bunching techniques associated with three general perspectives

to be specific: data pre-preparing, irregularity detection, and data projection/caution sifting. In the end, proposals for future inquires about took after by the conclusion are delineated toward the finish of this paper.

Authors[15] The objective of this paper is to give a review of a few works that utilize data mining techniques for intrusion detection and to address some specialized issues. They proposed another a thought in this paper will see intrusion detection from a data distribution center point of view and incorporate data mining and on-line expository handling (OLAP) for intrusion detection purposes. One of the significant impediments of the systems is that they need versatility to changing conduct designs. Some specialized issues were examined which are basic in building up a genuine versatile, ongoing intrusion detection system.

III. PROPOSED WORK

There are numerous current instruments for Intrusion detection system, yet the real issue is the security and exactness of the system. To enhance the issue of precision and the proficiency of the system an exceptionally regular arrangement approach i.e. decision tree is utilized. Proposed inquire about work acquaints a structure with build up a classifier in light of data mining techniques.

In this system ORNL dataset is given to Pre-preparing stages which arrange in j48 algorithm and diminish disrespectful highlights from the data set so data with less number of highlights will require to bolster to the classifier and will give effectiveness to the classifier. Machine learning apparatuses WEKA are utilized to examine the execution of datasets.

IV. EXPERIMENTAL SETUP

The exploratory approach followed in this examination incorporates ORNL dataset and grouping method i.e. J48 decision tree algorithm. The depictions of these philosophies are portrayed beneath. 4.1 WEKA , WEKA is an innovatory apparatus in the historical backdrop of the data mining and machine learning research groups. By putting endeavors since 1994 this device was produced by WEKA group. WEKA contains numerous inbuilt algorithms for data mining and machine learning. It is open source and uninhibitedly accessible stage free programming. The general population who are not having much information of data mining can likewise utilize this product effortlessly as it give adaptable offices to scripting tests. As new algorithms show up in explore International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014 15 writing, these are refreshed in programming. WEKA has likewise wound up one of the most loved device for data mining examination and advanced it by making numerous effective highlights accessible to all[1]. The means performed for data mining in WEKA are:

A. Preprocess the datasets.
   1) Load data
   2) Preprocess data

B. Analyze attributes.

Classify the datasets.

1) Select Test Options e.g:
   - ✓ Use Training Set
   - ✓ % Split,
   - ✓ Cross Validation
2) Run classifiers
3) View results

V. DATA DESCRIPTION

Oak Ridge National Laboratories (ORNL) have made 3 datasets which incorporate estimations identified with electric transmission system typical, unsettling influence, control, digital attack practices. Estimations in the dataset incorporate synchrophasor estimations and data logs from Snort, a reenacted control board, and transfers. The preparation dataset comprises of

4966 examples and contains 129 characteristics. The attacks composes are comprehensively classes into four gatherings-

Table 1. Types Of Attack

DoS	Denial of service
R2L	Remote to Local
U2R	User to Root
Probing	Surveillance, Port Scans, etc.

VI. J48 DECISION TREE

Arrangement is the way toward building a model of classes from a set of records that contain class names. Decision Tree Algorithm is to discover the way the properties vector carries on for various occasions. Likewise on the bases of the preparation occurrences the classes for the recently created cases are being discovered this algorithm produces the tenets for the forecast of the objective variable. With the assistance of tree order algorithm the basic conveyance of the data is effortlessly understandable [1]

J48 is an augmentation of ID3. The extra highlights of J48 territory meaning missing esteems, decision trees pruning, consistent trait esteem ranges, inference of principles, and so forth. In the WEKA data mining instrument, J48 is an open source Java usage of the C4.5 algorithm. The WEKA device furnishes various alternatives related with tree pruning. In the event of potential over fitting pruning can be utilized as a device for précising. In different algorithms the order is performed recursively till each and every leaf is unadulterated, that is the arrangement of the data ought to be as immaculate as would be prudent. This algorithm it produces the principles from which specific personality of that data is created. The goal is speculation of a decision tree until the point when it picks up balance off adaptability and exactness.

VII. RESULTS

For preparing the system, ORNL DATASET is considered which comprises of 4966 instances and contains 129 attributes.

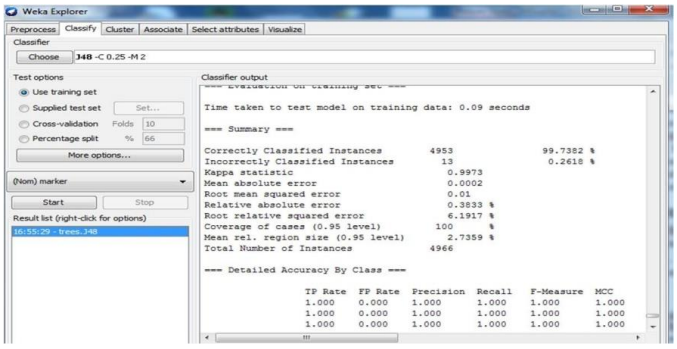

Figure 1. Shows result of dataset after training.

Once the system has been trained, it can be tested for its performance. The data sets include whole training set itself, 10 cross validation is applied on the training set, splitting the training dataset and providing a completely different test dataset. Based on the records of the different datasets results are obtained separately for the system as shown in the Table.

Table 2. Training The System By Cross Validation On Datasets

Datasets used for Testing	Correctly classified instances	Incorrectly classified instances	TP Rate	FP Rate	Precision	Recall	F-measure	ROC
7	92.509	7.409	0.075	0.056	0.313	0.075	0.071	0.579
9	92.586	7.4041	0.074	0.056	0.283	0.074	0.038	0.514
10	97.338	3.261	0.997	0	0.997	0	0	1

Table 3. Testing The System By Splitting Datasets On Different Percentage

Percentage split on training datasets	Correctly classified instances	Incorrectly classified instances	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC
50%	99.4443%	0.5557%	0.994	0.006	0.994	0.994	0.997	0.997
60%	99.4443%	0.5557%	0.994	0.006	0.994	0.994	0.994	0.996
66%	99.533%	0.467%	0.995	0.005	0.995	0.995	0.995	0.996
70%	99.484%	0.516%	0.995	0.005	0.995	0.995	0.995	0.998
80%	99.6626%	0.3374%	0.997	0.004	0.997	0.997	0.997	0.999

Table 1 shows the decision tree that is constructed after the system is trained. The number of leaves used to build the tree is 4848, and the size of the tree is 4877.

VIII. CONCLUSION

In this exploration we have actualized techniques for intrusion detection which gives better execution. In this examination we have researched in signature based intrusion detection which recognizes just known attacks. The future improvement of this system is, it expels its downside by executing a system that recognizes both obscure and known attack.ThisJ48 algorithm gave higher exactness over NB and SVM. This algorithm indicates 99.73% of precision.

IX. REFERENCES

[1]. Rebecca Bace and Peter Mell,"Intrusion Detection Systems", NIST Special Publication on Intrusion Detection Systems.

[2]. K.Nageswara rao, D.RajyaLakshmi, T.Venkateswara Rao," Robust Statistical Outlier based Feature Selection Technique for Network Intrusion Detection", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[3]. Ala' Yaseen Ibrahim Shakhatreh , Kamalrulnizam Abu Bakar ,"A Review of Clustering Techniques Based on Machine learning Approach in Intrusion Detection Systems", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011.

[4]. "NSL-KDD data set for network-based intrusion detection systems ", Available on: http://nsl.cs.unb.ca/NSL-KDD.

[5]. AdetunmbiA.Olusola,Adeola S.Oladele.,Daramola O.Abosede,"Analysis of KDD '99 Intrusion Detection Dataset for Selection of Relevance Features", Proceedings of the World Congress on Engineering and Computer Science 2010 Vol I WCECS 2010, October 20-22, 2010, San Francisco, USA.

[6]. M. Sathya Narayana1, B. V. V. S. Prasad2, A. Srividhya3 and K. Pandu Ranga Reddy", Data Mining Machine Learning Techniques – A Study on Abnormal Anomaly Detection System",International Journal of Computer Science and Telecommunications Volume 2, Issue 6, September 2011].

[7]. Mahmood Hossain," Data Mining Approaches For Intrusion Detection: Issues And Research Directions",Department of Computer Science, Mississippi State University, MS 39762, USA.

[8]. S Pandey, R Miri, S R Tandan "Diagnosis and Classification of Hypothyroid Disease Using Data Mining Techniques" International Journal of Engineering Research and Technology, (2013)

[9]. Khushboo Sharma, S R Tandan "An Optimized Parallel Confidence Measures Algorithm on Web Log Data" International Journal of Engineering Research and Technology (IJERT) (2013)

[10]. Asha Miri, S.R.Tandan, Rohit Miri "Pseudo Code to Eliminate Unwanted Data Sets for

Fuzzy Mining Association Rule" International Journal for Research in Applied Science & Engineering Technology (IJRASET) (2015).

[11]. Rohit Miri, Priyanka Tripathi, S R Tandan" Exploration of Novel Algorithm for Reduced Computational Time by Using Fuzzy Classification Technique in Data Mining" International Journal for Research in Applied Science & Engineering Technology (IJRASET) (2015)

[12]. Rohit Miri, Priyanka Tripathi, S R Tandan "Novel Algorithm For Finding The Range Of Fuzzy Values For Quantitative Data Sets By Data Mining And Fuzzy Technique" Journal Of Advanced Database Management & Systems Journal Of Advanced Database Management & Systems, (2015)

[13]. S R Tandan, Rohit Miri, ,Dr Priyanka Tripathi "A Bird Eye Review on Reduced Time Complexity by Using Data Mining and Fuzzy Techniques" international journal for research in applied science and engineering technology (ijraset), (2014)

[14]. S R Tandan Rohit Miri, Priyanka Tripathi "TRApriori Classification Based Algorithm by Fuzzy Techniques to Reduced Time Complexity" International Journal of Computer Science & Information Technology, International Journal of Computer Science & Information Technology, (2014)

[15]. Rohit Miri, Priyanka Tripathi, Keshri Verma, S.R. Tandan "Novel Algorithm to Reduced Computational Data Sets for Fuzzy Association Rule" International Journal For Research In Applied Science And Engineering Technology (IJRASET), (2015).

[16]. R Miri, P Tripathi, S R Tandan "Novel Algorithm for Reduced Computational Data by Using Fuzzy Classification and Data Mining Techniques" Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies" ACM, (2014)

[17]. Shoban Babu Sriramoju, "Review on Big Data and Mining Algorithm" in "International Journal for Research in Applied Science and Engineering Technology", Volume-5, Issue-XI, November 2017, 1238-1243 ISSN : 2321-9653], www.ijraset.com

[18]. Shoban Babu Sriramoju, "OPPORTUNITIES AND SECURITY IMPLICATIONS OF BIG DATA MINING" in "International Journal of Research in Science and Engineering", Vol 3, Issue 6, Nov-Dec 2017 ISSN : 2394-8299].

[19]. Ajmera Rajesh, Siripuri Kiran, " Anomaly Detection Using Data Mining Techniques in Social Networking" in "International Journal for Research in Applied Science and Engineering Technology", Volume-6, Issue-II, February 2018, 1268-1272 ISSN : 2321-9653], www.ijraset.com

[20]. Shoban Babu Sriramoju, Madan Kumar Chandran, "UP-Growth Algorithms for Knowledge Discovery from Transactional Databases" in "International Journal of Advanced Research in Computer Science and Software Engineering", Vol 4, Issue 2, February 2014 ISSN : 2277 128X]

[21]. SA Supriya. "A Survey Model of Big Data by Focusing on the Atmospheric Data Analysis." International Journal for Scientific Research and Development 5.10 (2017): 463-466.

[22]. Shoban Babu Sriramoju, Azmera Chandu Naik, N.Samba Siva Rao, "Predicting The Misusability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" , Vol V,Issue II,Febrauary 2014 ISSN : 2321-3469]

[23]. Ajay Babu Sriramoju, Dr. S. Shoban Babu, "Analysis on Image Compression Using Bit-Plane Separation Method" in "International Journal of Information Technology and

Management", Vol VII, Issue X, November 2014 ISSN : 2249-4510]

[24]. Monelli Ayyavaraiah, " A Study on Large-Scale Cross-Media Retrieval of Wikipedia Images towards Visual Query and Textual Expansion" in "International Journal for Research in Applied Science and Engineering Technology", Volume-6, Issue-II, February 2018, 1238-1243 ISSN : 2321-9653], www.ijraset.com

[25]. Shoban Babu Sriramoju, "Mining Big Sources Using Efficient Data Mining Algorithms" in "International Journal of Innovative Research in Computer and Communication Engineering" , Vol 2, Issue 1,January 2014 ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[26]. Siripuri Kiran, 'Decision Tree Analysis Tool with the Design Approach of Probability Density Function towards Uncertain Data Classification', International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X,Volume 4 Issue 2, pp.829-831, January-February 2018. URL : http://ijsrst.com/IJSRST1841198

[27]. Dr. Shoban Babu Sriramoju, "A Review on Processing Big Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol-2, Issue-1, January 2014 ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[28]. Amitha Supriya. "Implementation of Image Processing System using Big Data in the Cloud Environment." International Journal for Scientific Research and Development 5.10 (2017): 211-217.

[29]. Shoban Babu Sriramoju, Dr. Atul Kumar, "An Analysis around the study of Distributed Data Mining Method in the Grid Environment : Technique, Algorithms and Services" in "Journal of Advances in Science and Technology" Vol-IV, Issue No-VII, November 2012 ISSN : 2230-9659]

[30]. Monelli Ayyavaraiah, "Nomenclature of Opinion Miningand Related Benchmarking Tools" in "International Journal of Scientific & Engineering Research" Vol 7,Issue 8, February 2018, ISSN 2229-5518]

[31]. Shoban Babu Sriramoju, Dr. Atul Kumar, "An Analysis on Effective, Precise and Privacy Preserving Data Mining Association Rules with Partitioning on Distributed Databases" in "International Journal of Information Technology and management" Vol-III, Issue-I, August 2012 ISSN : 2249-4510]

[32]. Monelli Ayyavaraiah, "Review of Machine Learning based Sentiment Analysis on Social Web Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 4,Issue 6,March 2016 ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

[33]. Shoban Babu Sriramoju, Dr. Atul Kumar, "A Competent Strategy Regarding Relationship of Rule Mining on Distributed Database Algorithm" in "Journal of Advances in Science and Technology" Vol-II, Issue No-II, November 2011 ISSN : 2230-9659]

[34]. Shoban Babu Sriramoju, Dr. Atul Kumar, "Allocated Greater Order Organization of Rule Mining utilizing Information Produced Through Textual facts" in "International Journal of Information Technology and management", Vol-I, Issue-I, August 2011 ISSN : 2249-4510]

[35]. Guguloth Vijaya, A. Devaki, Dr. Shoban Babu Sriramoju, "A Framework for Solving Identity Disclosure Problem in Collaborative Data Publishing" in "International Journal of Research and Applications" (Apr-Jun 2015 Transactions), Vol 2, Issue 6, 292-295