ISSN: 2456 - 3935

International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 04, April 2017, pp. 09 – 13

Role of Decision Trees in Intrusion Detection Systems: A Survey

Pratham Harshit Rajmahanty
PG Scholar, School of Computing Science and Engineering, VIT University-Chennai Campus, Chennai, India
Email: prathamharshit0827@gmail.com

S. Ganapathy
Assistant Professor (Sr.), School of Computing Science and Engineering, VIT University-Chennai Campus, Chennai, India
Email: sganapathy@vit.ac.in

Abstract: *Recently, the number of intruders is increasing in computer networks gradually according to the internet user growth. We are in the position to introduce new techniques for detecting the intruders. For this purpose, many intrusion detection systems have been proposed in the past by various researchers in the past two decades. Even though, not yet achieved better detection accuracy in dynamic computer networks namely mobile ad-hoc networks and Internet due to the introduction of new intruders daily. This survey presents many decision tree algorithms which are proposed by various researchers in the past for effective decision making on intrusion detection systems and some other decision making systems. Moreover, a comparative analysis also made in this paper for demonstrating their capability. Finally, we suggest some ideas for enhancing the performance of the existing decision tree algorithms for improving the performance in terms of detection accuracy over the intrusion datasets.*

Keyword: *Decision Trees; Feature selection; Classification; Neural Networks; Intrusion Detection Systems.*

1. INTRODUCTION

As of late, web has turned into a develop need of everyday life. The present web based applications and master systems are defenseless because of the entry of new dangers that outcome in various sorts of harms prompting to imperative misfortunes. In this way, the significance of learning security is developing rapidly. The principal fundamental objective information of information of learning security is to create protective data system which are secure from unapproved get to, utilize, divulgence, intrusion, adjustment, or demolition. Additionally, the information security is minimizes the dangers including the three primary security objectives to be specific classification, uprightness and handiness.

Various systems are planned inside the past to spot and stop the Internet-based assaults. The first essential systems among them are Intrusion Detection Systems (IDS) since they oppose outer assaults adequately. In addition, IDSs offer a mass of resistance that beats the assault of PC systems on the web. IDS may be adjusted watch the diverse sorts of assaults on system correspondences and framework use wherever the standard firewall not ready to perform well. Intrusion detection is predicated on partner degree presumption that the conduct of gatecrashers takes issue from a legitimate client. For the most part, IDSs are extensively grouped into two classes in particular oddity and abuse discovery systems bolstered their location approaches. Peculiarity Intrusion detection figures out if or not deviations from the set up customary utilization examples are frequently hailed as intrusions. On the other hand, abuse perception systems are recognizes the infringement of authorizations adequately. Intrusion identification systems are regularly composed by wise specialists and arrangement procedures. Most IDSs include two areas particular decision preprocessing segment and intrusion detection stage. The intrusions known by the IDSs are frequently anticipated adequately by creating partner degree intrusion hindrance system. The main objective of this paper is to provide effective survey on decision tress which is contributing more for detecting intruders in intrusion detection systems.

The objective of IDS is to find noxious movement. To finish this, the IDS monitor all approaching and active movement. There are numerous methodologies on the execution of IDS. Among those, two are the most widely recognized:

Anomaly Detection: This strategy depends on the location of activity anomalies. The deviation of the observed activity from the conventional profile is measured. Shifted totally unique executions of this strategy are proposed, supported the measurements utilized for measuring movement profile deviation.

Misuse/Signature Detection: This strategy appearance for examples and marks of officially striking assaults inside the system activity. Continually upgraded data is now and again used to store the marks of remarkable assaults. The methods this strategy manages intrusion detection takes after the implies that hostile to anti-virus software works.

Intrusion Detection Systems (IDS) turned into a standard component in security foundations as they allow network administrators to find approach infringement. Current IDS have assortment of genuine downsides: Current IDS are infrequently tuned to distinguish striking administration level system assaults. This abandons them inclined to unique and novel malicious assaults. Information over-load: Another perspective that doesn't relate on to misuse detection however is exceptionally fundamental and what extent of information an expert will with proficient analysis. The amount of learning must review and looks forward rapidly. Contingent upon the intrusion identification instruments used by an association and its size there's the likelihood for logs to prevail in a great many records for every day. False positives: A run of the mill dissension is that the measure of false positives an IDS can create. A false positive happens once ordinary assault is mistakenly delegated vindictive and treated subsequently. False negatives: This can be the situation wherever an IDS doesn't create an alarm once an intrusion is genuinely happening. Information mining will enhance intrusion discovery by tending to every last one of the previously mentioned issues.

A decision tree is laid out as "a prognostic demonstrating procedure from the fields of machine learning and measurements that manufactures a direct tree-like structure to display the fundamental example". Decision trees square measure one case of a grouping algorithmic run the show. Grouping could be an information preparing strategy that doles out items to no less than one of numerous predefined classes. From partner intrusion location point of view, arrangement calculations will portray organize data as vindictive, checking, or alternate class of intrigue exploitation information like source/goal ports, data preparing addresses, furthermore the scope of bytes sent all through a connection.

Decision trees could be a viable tool within the intrusion detection toolkit; the technique has to satisfy a minimum set of necessities. The technique has to be useful to the intrusion analysis mission and turn out real results for a company. In addition, decision trees should be distinctive among existing tools. If alternative tools exist with a similar practicality provided by decision trees, then decision trees could also be redundant and redundant.

2. RELATED WORKS

There are many works have been proposed in the past by various researchers in this direction. Among them, Zehra Cataltepe et al (2016) [1] proposed a semi-regulated technique for network anomaly identification which utilizes online group, online element decision and decision tree as its building pieces. They contrasted and existing systems and infer that to the unattended anomaly identification and the semi regulated technique consolidates a far superior precision execution. Moreover, the element decision procedure allows each class to be weighted something else. Especially for network intrusion detection wherever keeping away from false negatives is to a great degree fundamental, it has inclination to trust that this capacity may well be useful. There are assortments of different future work headings.

Jamal Esmaily et al (2015) [2] proposed a method which is in light of Artificial Neural Networks and Decision Trees to plan a precise Intrusion Detection System with high recognition rate and low false alert rate. Their system is comprises of two distinct stages. The primary period of their calculation is to make a substitution dataset through feeding the arrangement consequences of the decision tree and multi-layer perceptron organize on dataset. In their second stage, the multi-layer perceptron is utilized once again to arrange the information inside the new dataset. This hybrid approach is useful to realize promising outcomes of the system. Moreover, it has low far and also promises the reliable leads for real world applications. Finally, they achieved better and potential results in their experiments.

Kathleen Goeschel (2016) [3] proposed a model which is the combination of support vector machines (SVM), decision tree algorithms and Bayesian classifiers. To start with, the SVM is prepared in view of another order strategy which depends on twofold grouping to the dataset for indicating whether the case is an assault or typical activity. In Second, assault activity is directed through a decision tree for grouping. Third, Naïve Bayes and the decision tree will then vote on any unclassified assaults. Neha and Dharmaraj [4] (2015) have directed an explored different avenues regarding two distinctive characterization calculations furthermore on two diverse datasets.

Their exploratory results are demonstrated that C4.5 with pruning result in higher precision. Shailendra Sahu and B M Mehtre (2015) [5] proposed the Kyoto 2006+ information set which is based on three year of genuine traffic data. It used J48 decision tree for network intrusion detection and also they achieved 97.23% of accuracy.

Shangguang Wang (2016) [6] proposed a new method which is in view of a self-selection decision tree, which can support the VHO among WAVE, WiMAX, and 3G cell. The decision tree settles on decision as per client inclinations, and the input decision strategy in accordance with the criticism of administrations and developments on vehicles can dodge the negative effect of administration changes and development changes. The technique mirrors the particular needs of vehicle to the system. Besides, there might be some other system characteristics and client inclinations that were not considered. Furthermore, the proposed decision tree strategy can be further improved. Abdul Hameed et al (2016) [7] introduced a new reference free and lightweight model for perceptual video quality prediction and it also based on the energy-efficient and content-aware FEC scheme. The quality of the prediction model can reveal the nonlinear relationships between perceptual quality and features which are related to video content, source coding parameters and network conditions.

Fangming Ye et al (2015) [8] developed a versatile board-level practical fault diagnosis technique which depends on incremental decision trees. It has demonstrated how flawed parts can be grouped in light of the discriminative capacity of the disorders in DT preparing. The finding system has been built as a binary tree with the most discriminative disorder as the root furthermore the last repair proposals are accessible as the leaf nodes of the tree. Along these lines, exact determination is completed easily with a little number of disorders. Moreover, they additionally joined the conclusion learning from debug technicians with the information picked up from coming up failing boards during the volume manufacturing. Hence, the issue of low conclusion precision toward the start of volume of production can be eased. Emmanuel Sapin et al (2015) [9] proposed an ant colony optimization (ACO) algorithm for the innovation of gene–gene interactions in GWAS data. They guaranteed that the ACO algorithm that doesn't fuse learned information or heuristic information oppositely to the methodologies. In refinement to the methodologies in, SNP connections are found by looking a full data set of GWAS information, involving a few SNPs and individuals.

Bin Zhou et al (2015) [10] proposed a novel hybrid methodology supported decision tree methodology and regeneration is projected. Six feature indices area unit designed for coaching the decision tree classifier to reason is landing mode and grid-connected mode. Safaa and Firas (2015) [11] proposed two levels of IDS system to research network traffic on completely different granularities. It's completely different from the on the market IDSs within which it adapts with network state of affairs once it's under fire or not. Its detection levels are coarse-grained IDS and fine-grained IDS. The foremost vital options for the two categories namely traditional and R2L which are heavily overlap that limit the detection rate of R2L attacks.

Ganapathy et al (2011) [12] proposed a new classification algorithm using C4.5 decision tree for effective classification. The same set of authors introduced the enhanced version of their algorithm with the help of intelligent agents (Ganapathy et al 2012) [13]. Jaisankar et al (2012) [14] proposed a new rough set based decision tree algorithm for enhancing the intrusion detection accuracy. They have used C 4.5 decision tree for decision making along with the fuzzy rough set theory. Sindhu et al (2012) [15] proposed another intrusion detection system which is lightweight IDS for multi-class arrangement. Firstly, the info example is reprocessed and repetitive occasions are evacuated. Next, a wrapper basically based feature selection algorithm is adjusted that elements a bigger effect on minimizing the strategy nature of the classifier. At long last, a neuro tree model is utilized in light of the fact that the characterization motor that granted a detection rate of 98.4% that is better than NN and expanded C4.5. Sannasi et al (2013) [16] made a survey on the intelligent intrusion detection systems for computer networks. The authors discussed many feature selection and classification techniques in their survey and also considered the contribution of the decision trees.

TABLE 1 THE PERFORMANCE OF THE DECISION TREE BASED INTRUSION DETECTION SYSTEMS.

S. No.	Author(s) & Year	Accuracy (%)
1	Shangguang Wang et al, 2014 [6]	60.54%
2	Abdul Hameed et al, 2016 [7]	88.89%
3	Fangming Ye et al, 2015 [8]	55.00%
4	Emmanuel Sapin et al, 2015 [9]	91.02%
5	Bin Zhou et al, 2015 [10]	88.90%
6	Kyoungok Kim (2016) [17]	79.23%
7	Zahra Cataltepe et al, 2016 [1]	96.28%
8	Jamal Esmaily et al, 2015 [2]	99.89%
9	Kathleen Goeschel, 2016 [3]	99.62%
10	Neha G. Relan et al, 2015 [4]	98.45%
11	Shailendra Sahu et al, 2015[5]	97.23%
12	Sannasi Ganapathy et al 2013 [16]	99.88%
13	Ganapathy et al 2012 [13]	99.84%
14	Ganapathy et al 2011 [12]	99.72%
15	Ganapathy et al 2011 [18]	99.32%

In spite of all the existing decision tree algorithms are contributing a lot for decision making. Eventhough, it must be improved with dynamic decision making nature on new environments. Recently, the decision tree algorithms are playing major role on decision making in all the intrusion detection systems for handling the new threats.

The comparative analysis is made based on their performance over the intrusion detection dataset. Many existing decision tree methodologies have been considered for comparative analysis. From that, Artificial Neural Networks based Decision Tree is performing well with 99.89% classification accuracy. The lowest performance has been achieved by Fangming Ye et al [8].

Table 1 shows the comparative analysis between various decision tree algorithms which are proposed in the past by various researchers in this direction. Among them, Jamal Esmaily et al [2] achieved 99.89% detection accuracy which is highest detection accuracy in the existing works.

3. SUGGESTION PROPOSED

Many decision tree algorithms have been proposed by various researchers for effective decision making. Noone achieved better detection accuracy over the new types of attacks which are occurring dynamically in networks. For achieving better classification accuracy through this decision tree algorithms can use the soft computing techniques. Moreover, we can use the intelligent agents for enhancing the performance.

4. CONCLUSION

An effective survey made with many decision tree algorithms which are proposed by many researchers in the past for effective intrusion detection and decision making over different the datasets. End of this survey, we conclude the contribution of the decision tree algorithms are very important for making suitable decisions over the datasets. Further works in this direction could be the proposal of new decision tree algorithms with soft computing techniques for effective decision making.

REFERENCES

[1] Zehra Cataltepe, Umit Ekmekci, Tanju Cataltepe, Ismail Kelebek,(2016) "Online Feature Selected Semi-Supervised Decision Trees for Network Intrusion Detection", *IEEE Network Operations and Management Symposium*, pp. 1085 - 1088.

[2] Jamal Esmaily, Reza Moradinezhad and Jamal Ghasemi, (2015), "Intrusion detection system based on Multi-Layer Perceptron Neural Networks and Decision Tree", *2015 7th Conference on Information and Knowledge Technology*, pp. 1 – 5.

[3] Kathleen Goeschel, (2016), "Reducing false positives in intrusion detection systems using data-mining techniques utilizing support vector machines, decision tree, and naïve Bayes for off-line analysis", *SouthEastCon*, pp. 1-6.

[4] Neha G. Relan and Dharmaraj R. Patil, (2015), "Implementation of network intrusion detection system using variant of decision tree algorithm", *2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE)*, pp.1-5.

[5] Shailendra Sahu and B M Mehtre, (2015), "Network Intrusion detection system using J48 Decision Tree", *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2023-2026.

[6] Shangguang Wang, Cunqun Fan, Ching-Hsien Hsu, Qibo Sun and Fangchun Yang, (2016), "A Vertical Handoff Method via Self-Selection Decision Tree for Internet of Vehicles", *IEEE Systems Journal*, Vol. 10, Issue. 3, pp. 1183-1192.

[7] Abdul Hameed, Rui Dai and Benjamin Balas, (2016), "A Decision-Tree Based Perceptual Video Quality Prediction Model and its application in FEC for Wireless Multimedia Communications", *IEEE Transaction on Multimedia*, Vol. 18, Issue. 4, pp. 764-774.

[8] Fangming Ye, Zhaobo Zhang, Krishnendu Chakrabarty and Xinli Gu, (2016), "Adaptive Board- Level Functional Fault Diagnosis using Incremental Decision Trees", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 35, Issue. 2, pp. 323-326.

[9] Emmanuel Sapin, Ed Keedwell and Tim Frayling, (2015), "Ant colony optimisation of decision tree and contingency table models for the discovery of gene-gene interactions", *IET Systems Biology*, Vol. 9, Issue. 6, pp. 218-225.

[10] Bin Zhou, Chi Cao, Canbing Li, Yijia Cao, Chen Chen, Yong Li and Long Zeng, (2015), "Hybrid islanding detection method based on decision tree and positive feedback for distributed generations", *IET Generation, Transmission & Distribution*, Vol. 9, Issue.14, pp.1819-1825.

[11] Safaa O. Al-mamory and Firas S.Jassim, (2015), "On the designing of two grains levels network intrusion detection system", *Karbala International Journal of Modern Science*, Vol.1, No.1, pp. 15-25.

[12] Ganapathy S, Yogesh P, Kannan A, (2011), "An intelligent intrusion detection system for mobile ad-hoc networks using classification techniques", *Computers and Communication Information Systems (CCIS)*, pp. 117-122.

[13] Ganapathy S, Yogesh P, Kannan A, (2012), "Intelligent agent-based intrusion detection system using enhanced multiclass SVM", *Computational intelligence and neuroscience*, Vol. 2012, pp. 1-9.

[14] Jaisankar N, Ganapathy S, Yogesh P, Kannan A, (2012), "Intelligent intrusion detection system using fuzzy rough set based C4. 5 algorithm", *International ACM Conference proceedings on Advances in Computing, Communications and Informatics*, pp. 596-601.

[15] Sindhu S, Geetha S, Kannan A, (2012), "Decision Tree based light weight intrusion detection using a wrapper approach", Expert Systems with Applications, Vol.39, Issue.1, pp. 129-141.

[16] Sannasi Ganapathy, Kulothungan Kanagasabai, Muthurajkumar Sannasy, Muthusamy Vijayalakshmi, Palanichamy Yogesh, Arputharaj Kannan, (2013), "Intelligent feature selection and classification techniques for intrusion detection in networks: a survey", *EURASIP Journal on Wireless Communications and Networking*, Vol. 271, Issue.1, pp. 1-16.

[17] Kyoungok Kim, (2016), "A Hybrid Classification Algorithm by subspace partitioning through Semi-Supervised Decision Tree", *Pattern Recognition*, Vol. 60, Issue. C, pp. 157-163.

[18] Ganapathy S, Rajesh Kambattan K, Veerapandian N, Pasupathy M, (2011) "An Intelligent Intrusion Detection Model for MANET's based on Hybrid Feature Selection", *Artificial Intelligent Systems and Machine Learning*, Vol. 3, Issue.13, pp. 849-852.

Authors Biography

Pratham Harshit Rajmahanty, is a PG Student in the School of Computing Science and Engineering in VIT University, Chennai Campus, Chennai, Tamilnadu, India. He received his B.C.A degree from PRSU, Raipur, India. His research interests are computer networks, Data Mining and security.

S.Ganapathy, is working as an Assistant Professor (Sr.), School of Computing Science and Engineering in VIT University, Chennai Campus, Chennai, Tamilnadu, India. He has completed his M.E and Ph.D degrees in Computer Science and Engineering at Anna University, Chennai. He has published more than 40 journal and conference papers in reputed publishers such as IEEE, ACM, Elsevier, Springer, IOS Press, IGI Global etc. His research interests are computer networks, Data Mining and security.