Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 4, 2020

Review Article

INTRUSION DETECTION SYSTEM USING MODIFIED J48 DECISION TREE ALGORITHM

D. Parameswari[1], Dr.V. Khanaa[2]

[1]Research Scholar, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India. parameswariridu@gmail.com
[2]Dean-Info, Department of Information Technology, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India. drvkannan62@yahoo.com

Received: 03.12.2019 *Revised: 06.01.2020* *Accepted: 08.02.2020*

Abstract
We have been used most popular J48 algorithm to evaluate the data. This algorithm is mainly used to evaluate different applications and carry out the exact results over the evaluation process. The main goal of developing this modified J48 choice tree algorithm is to reduce the quest manner in compare with the modern-day lively listing. We have implemented the MAC cope with of the device into CADL. To become aware of an outsider they have been used changed J48 choice tree set of rules. This algorithm produces the similar result as GA.

Keywords: Genetic Algorithm, MAC, CADL, J48 Decision Tree Algorithm.

DOI: http://dx.doi.org/10.31838/jcr.07.04.135

INTRODUCTION

Using the updated J48 selection tree set of guidelines, the safety o f the protocol tool identity is achieved after the 64byte protocol s tructure standardization and mutation capabilities of genetic tec hnique and go over. Achievement of the selection tree algorithm and the diagnosed effects are mentioned on this bankruptcy.
The primary objective of creating this modified set of regulations for the J48 option tree is to decrease the search procedure in review with the existing active listing. The tool's MAC deal is available under CADL. This listing is used as an input for detecting the intruder. On the way to selecting, it is proposed to add a modified set of regulations in the J48 option tree. That collection of rules is similar to

Precondition

All the current lively directory list elements are taken care of in an ascending order and generated list with specific device illustration of the community.

Step 1: The Current Active Directory List (CADL) Has been used to generate the evaluation of MAC value.
Step 2: The present day energetic directory record elements are normal as much a tree based totally over the price over the present MAC cope together with on the devices
Step3: At every node left plant made together with much less rate elements yet excellent plant wrought with greater virtue factors.
Step 4: The foot 3 are observed till whole the elements between the CADL is blanketed among the tree. Stability
Step 5: Estimate the Left Significant Byte (LSB) to identify the gain (Similarity) of the Device of the observed and suspected packet generated value.
Step 6: If partial bits are amount since returns 1 otherwise rejoinder zero namely a develop up.. durability
Step 7: calculate the expanded value
Step8: If the competencies maximum worth is internal the CADL then to that amount device is authenticated gadget over and above propose for the suspected device.

This manner is carried out as share on the data mining according to beat atop the IDS.

DATA MINING APPROACHES FOR IDS DECISION

The process of extracting patterns from data is Statistics mining. Data mining is a vital method with the resource from which current business organization turns knowledge into information obtained by business agency intelligence. It is commonly used in a wide variety of profiling activities at the moment. Help within th is the most significant explanation for the use of statistics mining Evaluation of set and behavioral analysis. The generation of mining information is advanced for the processing of large data portions and the discovery of secret and unrecognized information. Mining records typically include four clustering, form, regression and association training sessions. The assignment of clustering is to cross "comparable" companies and systems within the details, without the use of the reg

Decision Tree Algorithm

An expansion tree plays the classification of a given pattern of information through severe pick ranges to help us acquire a complete desire to close. This kind of pick series is depicted in a tree-shaped form. For the classification of uncertain statistics the tree form is used. A preference tree with a number of discrete (symbolic) labels of magnificence is referred to as a category tree, also as a preferential one.

IDE3 Algorithm

The set of regulations for the ide3 (iterative dichotomiser 3) tree of choice was introduced in 1986 with the aid of quinlan ross[3][7]. It's miles focused largely on the set of rules for hunting, and is done in sequence. Like distinct desire tree algorithms the tree is designed in steps; tree boom and pruning tree. Statistics are sorted in the tree building section at each node, in order to pick the consequences even as there may be too-a good buy noise or data within the schooling information set; therefore an intensive pre-processing of facts is achieved in advance than building a choice tree model with the ide3.

CART Algorithm

The cart (class and regression trees) is provided via breiman[5]. It creates bushes together with entire splendor or regression. The improvement about astonishing tree with the aid of cart is based entirely over the double splitting on attributes. It is also

based totally chiefly of the hunt version on Decision Tree development, yet can lie implemented in series [5]. Permanency Uses the splitting gini index test when choosing the spl The cart is not the same as other mainly based hunting algorithms because it is often used to test regression using wooden regression. The regression evaluation function is used to forecast a dependent variable (give up end result), provided the difficult and speedy predictor variables over a given time frame. Uses a lot of unmarried splitting vector norms like gi

Fuzzy Logic

It techniques the entry of group data, and explains steps that can be important for identification of anomalies. Fuzzy good judgment [8] is a much appreciated, not exceptional, type of feeling. With rationale it gives, which is approximate in desire to be set and real. In the evaluation of the conventional principle of good judgment, in which binary sets have -valued the correct judgment: (real or false), mystical correlations robotics. A murky device [9] consists of a employ of linguistic statements in particular based on expert know-how. This facts is typically of the form of if-then policies. A litigation then An destination can keep top notch by way of course about making use of a set on murky common sense rules, based totally concerning the attributes' linguistic values. A assessment concerning the a variety of decision arbor algorithms is introduced below within Table 2.5.

Table 1.1: Comparison of various decision tree algorithms

S.no	Classifier	Method	Parameters	Advantages	Disadvantages
1	Decision Tree [91][92]	Decision tree is based on the binary classification tree.	Positive and negative instances.	1. Construction does not require any domain knowledge. 2. Can handle high dimensional data. 3. Able to process both numerical and categorical data.	1. Output attribute must be categorical. 2. Limited to one output attribute. 3. Decision tree algorithms are unstable. 4. Trees created from numeric datasets can be complex.
2	Iterative Dichotomiser 3 [IDE3][81][80]	Data is sorted at every node during the tree building phase, in-order to select the best splitting single attribute	It only accepts categorical attributes in building a tree model	1.Easy to detect in the sorted elements	1.Does not give accurate results when there is too-much noise or details in the training data set
3	C4.5[79]	It uses the gain ratio impurity method to evaluate the splitting attribute	Accepts both continuous and categorical attributes in building the decision tree.	1. By replacing the internal node with a leaf node, thereby reducing the error rate	1.Unable to detect the unsorted non categorical data set 2.Noise data require pre processing for the detection
4	Classification and regression trees (CART)[15]	CART is based on the binary splitting of the attributes.	Uses both numeric and categorical attributes for building the decision tree	1.In-built features that deal with missing attributes	1.Data is sorted at every node to determine the best splitting point. 2.The linear combination splitting criteria is used during the regression analysis.
5	Support Vector Machine [42]	Support vector machine	The effectiveness of SVM lies in the selection of kernel and soft margin parameters.	1.Highly Accurate 2. Able to model complex nonlinear decision boundaries	1. High algorithmic complexity and extensive memory. 2. The choice of the kernel is difficult
6	Fuzzy[38]	Fuzzy logic helps to smooth the abrupt separation of normality and abnormality.	Range values of zero and one.	1. It was able to detect non malicious port scans launched against the system from the local domain.	1. Unable to detect a greater diversity of intrusions.
7	Modified J48	Range determination for the communication device based on Sequential value.	Ranges are determined based on the value and the minimal attributes.	1. Reduce the search time for the Sorted elements. 2. The value is regenerated for each iteration with minimal tree approach.	1. The search required the sorting as preprocess.

These algorithms limitations are overcome into the implementation about the genetic strategy and the highlights are mentioned beyond comparable lookup work.

MODIFIED J48 DECISION TREE ALGORITHM

The 16-bit example of the mac cope method is given in the new energetic directory list. The modified j48 choice tree set of rules discusses the uniform facts that support the results from

choosing a characteristic to break the statistics. The role with the best uniform record value shall be used to make the pick. Then the collection of regulations on the smaller s resumes

Then inside the tree of choice a leaf node is generated telling you to choose the elegance. In this case, the modified j48 desire tree set of rules produces a selection node that uses the anticipated class fee better within the tree. If the created lsb fee for cadl and incoming protocol system mac is the same, the system recommended for intrusion will be authenticated otherwise.

The form of the preference tree changed by j48 is shown in Discern 5.1. The tree's principal stage is a single header node. It's far from being just a pointer node to its babies. The tree's second one stage has 2 subtrees from 1 to 2 graded.

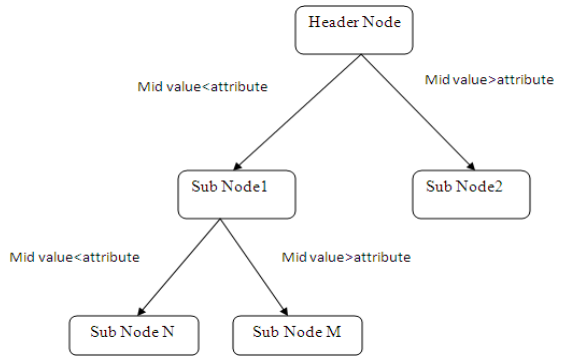

Figure 1.1: Structure of the modified J48 decision tree

Modified J48 algorithm Pseudo Code

The following pseudo code is used to build decision trees

1. Test [Initial User List from Current Active Directory List]
2. For base instances. The standardized statistics benefit from separating the 16 bit system from the least giant bit. Let's better be the attribute with the highest standardized gain of information {Allowed to interact on the network}
3. Creating a collection node that divides the least nice bits or the incredible bit for going over four on a fine choice.
4. Recourse at the sublists near via splitting regarding a maximum, then accumulate those nodes so node boys The coming charge beside the MAC law transforms the usage over IANA distribute or compares the values. The C++ coding is aged after sue the amended J48 decision tree.

As described in step1and step2 the generated mac deal with values is exceeded as a parameter to the changed set of rules in the j48 preference tree. The manner in which the pattern is applied is specified by using manufacturer and system details 08-00-06-04-00-01.

The 08h fee is transferred in parameter form. The 08h search process is presented with a cadl that is to be tested. The algorithm of the modified j48 preference tree

Assign CADL is a root of the tree for the search process

Tree = {Elements of CADL} // where CADL consists of SMA, SDA, DMA, DDA

Procedure for Modified J48 Algorithm Calculation

Assign L as a first element // L is a low range value
Assign H as Last element // H is a Highest value
D = | L- H | // Difference
M = loc (D/2) //M- Selection of Middle Element Search process
If (M = e) // e – search element
Declare element found
Else
If (M < e)
subtree = Tree(L) …. Tree (M);
Tree = subtree;
Else
subtree = Tree(M) … Tree(H) ;
Tree = subtree;
End if
End if

EXPERIMENTAL RESULT

The method over intrusion detection is successful together with real-time packets which could stay accrued beside protocol observations or converted among a record. This collection is awesome from the regular communication functions, and from the attack functions. In the hybrid educational group with 800 nodes, 3 varieties of protocol structure (arp, snmp-alg, and icmp) are analyzed.

ARP_IDS=$\sum_{i=1}^{N}(idARP)/Initiated\_intrusion$ (1)

Where IdARP in ARP Protocol is Considered Intruders. The efficiency of the IDS mechanism is dependent on the packet identifier. After evaluations of the supplier and system the number of packets is allowed to communicate in the authenticated manner. Table 1.2 shows the outcome of packet intrusion detection for the ARP protocol.

Table 1.2: Result of the ARP protocol packet Intrusion detection (10 files)

S. No	Number of packets	Initiated Intrusion	Number of ARP packets	ARP Identified	% of ARP Identified Vs Total	% of ARP Identified Vs Initiated	%of Identified ARP Vs ARP
1	52489	6650	28010	1525	2.91	22.93	5.44
2	52489	4107	28010	1271	2.42	30.95	4.54
3	49676	3881	23424	549	1.11	14.15	2.34
4	49676	5252	23424	891	1.79	16.96	3.80
5	51305	8606	25252	2244	4.37	26.07	8.89
6	51305	7478	25252	1766	3.44	23.62	6.99
7	49781	3779	26489	884	1.78	23.39	3.34
8	57756	3728	27169	824	1.43	22.10	3.03
9	63383	6869	24464	848	1.34	12.35	3.47
10	54761	4478	27906	1072	1.96	23.94	3.84
Min	49676.00	3728.00	23424	549.00	1.11	12.35	2.34
Max	63383.00	8606.00	28010	2244.00	4.37	30.95	8.89

Avr	53262.10	5482.80	25940	1187.40	2.25	21.65	4.57

The intrusion detection level in SNMP–ALG protocol against the initiated intrusion is given in the Eqn (2).

SNMP–ALG_ IDS=

$$\sum_{i=1}^{N}(idSNMP-ALG)/initiated\_intrusion \quad (2)$$

Where Id SNMP–ALG is a Identified Intruders in SNMP–ALG protocol. The result summary of the SNMP–ALG protocol packet intrusion detection is presented in Table 1.3, given below.

Table 1.3: Result of SNMP-ALG protocol packet Intrusion detection (10files)

S. No	Number of packets	Initiated Intrusion	SNMP-ALG packets	SNMP-ALG Identified	% of SNMP-ALG Identified Vs Total	% of SNMP-ALG Identified Vs Initiated	% of Identified SNMP-ALG Vs SNMP-ALG
1	52489	6650	3154	203	0.39	3.05	6.44
2	52489	4107	696	19	0.04	0.46	2.73
3	49676	3881	2583	98	0.20	2.53	3.79
4	49676	5252	932	45	0.09	0.86	4.83
5	51305	8606	720	44	0.09	0.51	6.11
6	51305	7478	746	51	0.10	0.68	6.84
7	49781	3779	4260	139	0.28	3.68	3.26
8	57756	3728	1080	25	0.04	0.67	2.31
9	63383	6869	2325	98	0.15	1.43	4.22
10	54761	4478	2989	117	0.21	2.61	3.91
Min	49676.00	3728.00	696	19.00	0.04	0.46	2.31
Max	63383.00	8606.00	4260	203.00	0.39	3.68	6.84
Avr	53262.10	5482.80	2478	83.90	0.16	1.65	4.44

The intrusion detection level in ICMP protocol against the initiated intrusion is given in the Eqn (3).

ICMP_ IDS=$\sum_{i=1}^{N}(idICMP)/initiated\_intrusion$ (3)

Where IdICMP is an identified intruders in ICMP protocol. The result summary of the ICMP protocol packet intrusion detection is presented in Table 1.4, given below.

Table 1.4: Result of ICMP protocol packet Intrusion detection (10files)

S. No	Number of packets	Initiated Intrusion	ICMP packets	Identified ICMP Packets	% of ICMP Identified Vs Total	% of ICMP Identified Vs Initiated	% of Identified ICMP Vs ICMP
1	52489	6650	6061	529	1.01	7.95	8.73
2	52489	4107	4033	113	0.22	2.75	2.80
3	49676	3881	4799	168	0.34	4.33	3.50
4	49676	5252	3189	128	0.26	2.44	4.01
5	51305	8606	2643	168	0.33	1.95	6.36
6	51305	7478	4732	317	0.62	4.24	6.70
7	49781	3779	4822	241	0.48	6.38	5.00
8	57756	3728	6964	184	0.32	4.94	2.64
9	63383	6869	7759	302	0.48	4.40	3.89
10	54761	4478	6645	184	0.34	4.11	2.77
Min	49676.00	3728.00	2643.00	113.00	0.22	1.95	2.64
Max	63383.00	8606.00	7759.00	529.00	1.01	7.95	8.73
Avr	53262.10	5482.80	5164.70	233.40	0.44	4.35	4.64

In line with the above evaluation in Tables 1.2, 1. Three and 1.4, the ARP protocol identified and detected 2,25 per cent of the total packets. It executed 21.65 percent of the entire intrusion initiated and 5.5 percent of the ARP protocol packets on the same time. In the general packets fetched from the com, the SNMP-ALG protocol recognized and detected far less variety of packets. The result comparison of IDS (10 files) in ARP, SNMP-ALG and ICMP protocols with initiated intrusion is depicted graphically in Figure 1.2.

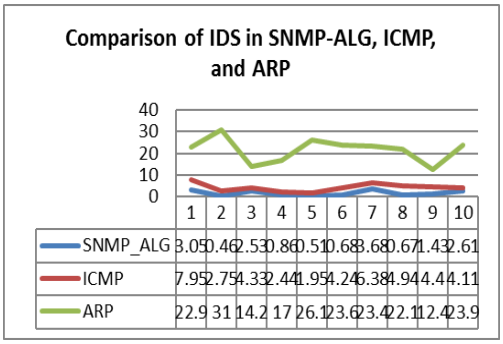

Figure 1.2: Comparison of IDS in SNMP-ALG, ICMP, and ARP (10 files) with initiated intrusion

CONCLUSION

In this bankruptcy the implementation portion of the modified J48 option tree set of rules and hunt system evaluation will be addressed. The looking time assessment of various network capacities (10 nodes, one hundred nodes, 500 nodes, 600 nodes, 800 nodes) is listed using the GA and the modified J48 set of rules. Implementation of the mechanism and its consequences from declaration of problem.

REFERENCES

1. Dewan Md, Farid, Mohammed Zahidur Rahman. "Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm", Journal of Computers, Vol 5, pp-23-31, Jan 2010, DOI:10.4.304/jcp 5.1.
2. Steinberg, D., and P. Colla. 1995. "CART: Tree-structured non-parametric data analysis". San Diego, Calif., U.S.A.: Salford Systems.
3. Quinlan, J.R. (1985b). "Decision trees and multi-valued attributes". In J.E. Hayes & D. Michie (Eds.), Machine intelligence 11. Oxford University Press.
4. Angham G. Hadi , Khudheir Jawad , Dina S. Ahmed , Emad Yousif. "Synthesis and Biological Activities of Organotin (IV) Carboxylates: A Review." Systematic Reviews in Pharmacy 10.1 (2019), 26-31. Print. doi:10.5530/srp.2019.1.5
5. Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). "Classification and Regression Trees", Wadsworth, Belmont, CA.
6. Steinberg, D., P. Colla, and K. Martin. 1998. CART—Classification and regression trees: Supplementary manual for Windows. San Diego, Calif., U.S.A.: Salford Systems.
7. Quinlan, J.R. (1986). "Induction of decision trees. Machine learning" 1, 81-106.
8. Yang, G., Lucas, R., Caldwell, R., Yao, L., Romero, M., Caldwell, R.Novel mechanisms of endothelial dysfunction in diabetes(2010) Journal of Cardiovascular Disease Research, 1 (2), pp. 59-63. DOI: 10.4103/0975-3583.64432
9. Saniee M., Habibi J., Lucas C. "Intrusion detection using a fuzzy genetics-based learning algorithm". Journal of Network and Computer Applications, 30(1), pp. 414 – 428. January 2007.
10. W. Li. (2004) "Using Genetic Algorithm for network Intrusion Detection", Proceedings of the United States Department of Energy Cyber Security, USA.