A Heuristic Approach for Efficient Detection of Intrusion

Naveen Mohan Prajapati
Patel College of Sc. & Tech.
Ralamandal, Indore
Madhya Pradesh, India

Atish Mishra
Patel College of Sc. & Tech.
Ralamandal, Indore
Madhya Pradesh, India

Praveen Bhanodia
Patel College of Sc. & Tech.
Ralamandal, Indore
Madhya Pradesh, India

ABSTRACT
The heuristic approach for efficient detection of intrusion is been proposed on this paper. The proposed framework uses new data preprocessing and filtration criteria which is data discretization to improve results of intrusion detection. It is more Accurate in comparison the existing methods. An overview of intrusion detection system is been presented. Also the present approaches for intrusion detection system are been described.

Keywords
ID, MDLP, ID3, KDD CUP 99.

1. INTRODUCTION
In present scenario, everyone in this huge world is using Internet to communicate with each other. Internet is now not only centered to the web mail or chat but also extended to the field of education, business, media and many more. Day by day, people are becoming more and more dependent to the Internet, which is making people's life easier. It is changing way of communication, business mode and even way of daily living. Now question is whether it is safe to deal each and everything using Internet? Is it secure enough to use? So the answer is 'no'. It is so because, as Internet grows, number of attackers & attacks also increase in a drastic manner. To make the use of internet much safer Intrusion detection concept was introduced by James Anderson long back in 1980[5], which defines that an intrusion attempt or threat to be potential possibility of a deliberate unauthorized attempt to access information or manipulate or render a system unreliable or unusable. Sites moved for using data mining in content of NIDS in the late of 1990's. After that researchers understood the need of standardized dataset to train IDS tool and developed Minnesota Intrusion Detection System (MINDS). Minnesota Intrusion Detection System (MINDS) combines signature based tool with data mining techniques, where Signature based tools (Snort) are used for misuse detection & data mining for anomaly detection.

2. INTRUSION DETECTION TECHNIQUES
All intrusion detection system use one of the two detection techniques [3]: signature based and statistical anomaly based

a) Signature/Misuse based IDS

b) Statistical/ Anomaly based IDS

A. Signature/Misuse Based IDS[2,3]

The signature based IDSs are also known as misuse detection which look for a specific signature to match for signaling an instruction with the provided signatures or patterns, but they are not as effective for unknown attack methods. Most popular intrusion detection falls in to this category. This means that an IDS using misuse detection will only detect known attacks.

B. Statistical/ Anomaly Based IDS[5,6]

Another approach of intrusion detection is called anomaly detection. Anomaly detection is usually applied for intrusion detection and computer security. It has been an active area of research since it was originally proposed by Denning 1987. Anomaly detection algorithms have the advantage that they can detect new types of intrusion. In this problem a set of normal data to train from and a new piece of test data is given to us. The goal of the intrusion detection system is to determine whether the test data belong to "normal" or to an "anomalous" behavior. However anomaly detection scheme suffers from a high rate of false positive and false negative. This occurs primarily because previously unseen system behavior are also recognized as anomaly, and hence flagged as potential instructions.

3. OBJECTIVE
The objective is to classify the information of a flow available in the form of 42 attributes (i.e. Network-based IDS) are normal or attack. This task of classification takes a lot of computation in model generation due to large data size.

The classification accuracy of [7] can be improved by using data preprocessing & applying data filtration with data discretization. The required time for the model generation can be reduced by applying supervised data discretization. This work will evaluate performance of hybrid approach of the classification over the Knowledge Discovery and Data Mining 1999(KDD'99) dataset.

4. PROPOSED SOLUTION
In decision tree classifiers, the criteria used for the attribute selection is as follows: First information gain of each attribute is computed then the attribute having maximum information gain is chosen. This means that an attribute with maximum values is chosen for splitting the tree. But in most of the cases, it is not necessary that an attribute with maximum values will be the best. Also ID3[17] algorithm uses the concept of information gain for selecting an attribute. The information gain is based on the concept of the probability. Probability based method is suitable for stochastic problems. But it cannot be the common criteria for attribute selection.

For solving this problem, a more accurate decision tree based classifier is been proposed in this paper. The proposed solution will use a new framework which reduce time to build model for intrusion detection also provide more accurate result by applying filtration of the data and then apply the classification.

5. PROPOSED METHOD

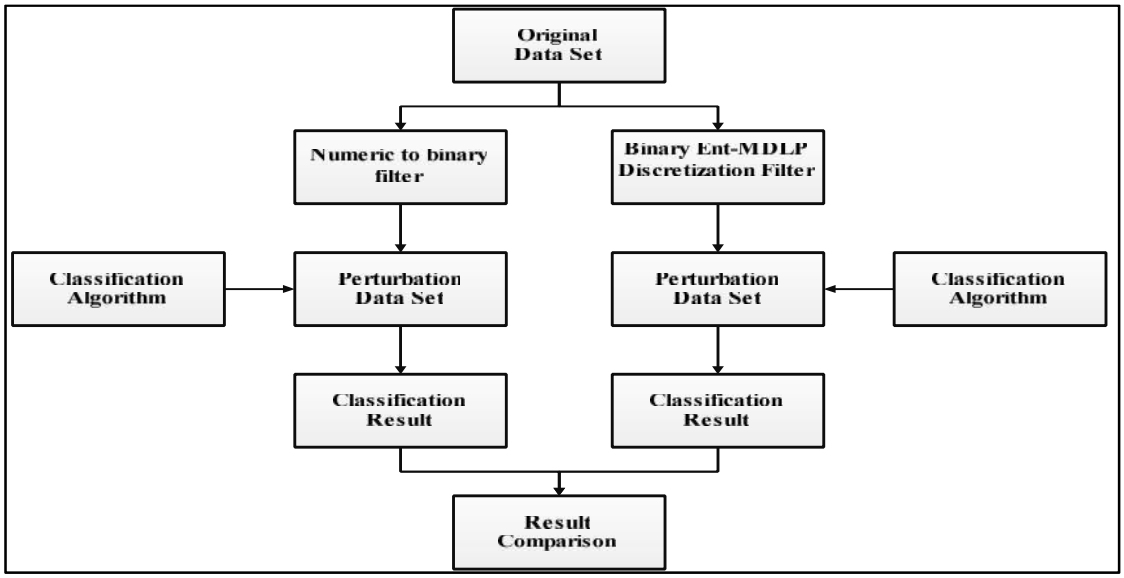

Figure 1. Framework of proposed solution

INPUT:

1. A training data set with class labels
2. List of attributes
3. Data filtration approach
4. Feature selection criteria.(used for splitting)

OUTPUT:

A Decision Tree

6.1 Procedure

1. Data Perturbation using:

a. Numeric to binary unsupervised filter in KDD99 Dataset and forward for classification.

b. Supervised Discretization to KDD99 Dataset and forward for classification.

2. Apply ID3 Algorithm [17] for preparation of Decision tree to both perturbation dataset.

6.2 Discretization

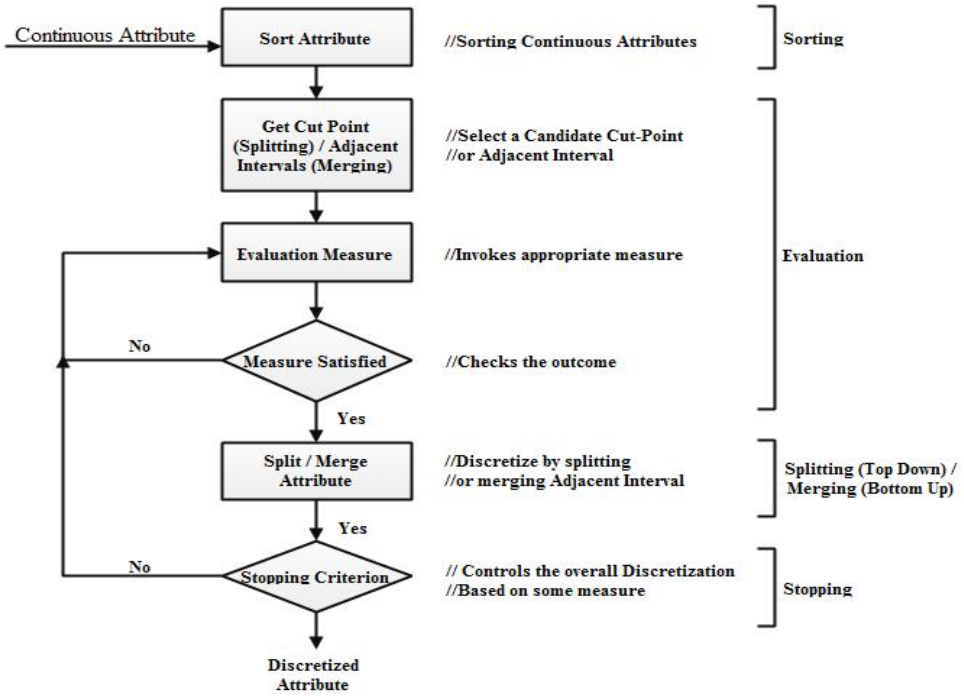

Figure 2. Process of Discretization

Ent-MDLP [18, 19] uses entropy measure from information theory to find a cut-point to split a range of continuous values into two intervals. Class information entropy is a measure of purity and it measures the amount of information which would be needed to specify to which class an instance belongs. The entropy measure is defined as:

$$E_e = E_1 + E_2$$

$$E_e = -p_{left}\sum_{i=1}^{k} p_{i,left}\log p_{i,left} - p_{right}\sum_{i=1}^{k} p_{i,right}\log p_{i,right}$$

Where

Ee = entropy of the cut-point

E1 = entropy to the left of the cut-point

E2 = entropy to the right of the cut-point

k = total number of classes

i = a practical class

pleft = number of instances to the left of cut-point / total number of instances, N

pright = number of instances to the right of cut-point / total number of instances, N

pi,left = num of instances of class i to the left of cut-point / number of instances to the left of cut-point

pi,right = {num of instances of class i to the right of cut-point } / { number of instances to the right of cut-point}

It considers one big interval containing all known values of a feature and then recursively partitions this interval into smaller subintervals until the stopping criterion satisfies. The stopping criterion was based on the MDL (Minimum Description Length) [18, 19] principle which is defined as:

$$Gain > \frac{\log_2(N-1)}{N} + \frac{\log_2(3^k-2) - k\,E + K_1E_1 + K_2E_2}{N}$$

Where:

$$E = -\sum_{i=1}^{k} p_i \log p_i$$

$$p_i = \frac{Number\ of\ instance\ of\ class\ i}{N}$$

gain = E – E_e = information gained by splitting at the cut-point

N = total number of instances in the attribute value list at each recursion

K_1 = number of classes to the left of the cut-point

K_2 = number of classes to the right of the cut-point

Step 1 of Proposed Scheme : Discretization algorithm

Input: A, continuous attributes.

C, class values in training set.

Algorithm:

For each attribute perform {

for each class perform{

find the no. of occurrences for each distinct value

find the max occurrence value in the class and consider it as key value

}

Sort the values of each attribute

By considering the key values as initial cut-points, form the subsets

for each subset perform{

Apply Ent-MDLP [18, 19] criteria and find the final cut-points

}

}

Output*:* Interval values of continuous attributes For Classification

Step 2 of Proposed Scheme : Classification using ID3 **[17]**
Algorithm:

Input: Perturbation Data Set

Algorithm:[17]

ID3 (Examples, Target Attribute, Attributes)

Create a root node for the tree

If all examples are positive, Return the single-node tree Root, with label = TRUE.

If all examples are negative, Return the single-node tree Root, with label = FALSE.

If number of predicting attributes is empty, then Returns the single node tree Root,

With label = most common value of the target attribute in the examples.

Otherwise Begin

A ← The Attribute that best classifies examples.

Decision Tree attribute for Root = A.
For each possible value, v_i, of A,

Add a new tree branch below Root, corresponding to the test A = v_i.

Let Examples(v_i) be the subset of examples that have the value v_i for A

If Examples(v_i) is empty

Then below this new branch add a leaf node with label = most common target value in the examples

Else below this new branch add the subtree ID3 (Examples(v_i), Target Attribute, Attributes – {A})

End

Return Root

Output: Decision Tree

Entropy

Entropy H(S) is a measure of the amount of uncertainty in the (data) set S (i.e. entropy characterizes the (data) set S).

$$H(S) = -\sum_{x \in X} p(x) \log_2 p(x)$$

Where,

S - The current (data) set for which entropy is being calculated (changes every iteration of the ID3 algorithm [17])

X - Set of classes in S

p(x) - The proportion of the number of elements in class x to the number of elements in set S

When H(S) = 0, the set S is perfectly classified (i.e. all elements in S are of the same class).

In ID3, entropy is calculated for each remaining attribute. The attribute with the smallest entropy is used to split the set S on this iteration. The higher the entropy, the higher the potential to improve the classification here.

Information Gain

Information gain IG(A) is measure of the difference in entropy from before to after the set S is split on an attribute A. In other words, how much uncertainty in S was reduced after splitting set S on attribute A.

$$IG(A) = H(S) - \sum_{t \in T} p(t) H(t)$$

Where,

H(S) - Entropy of set S

T - The subsets created from splitting set S by attribute A such that

$$S = \bigcup_{t \in T} t$$

p(t) - The proportion of the number of elements in t to the number of elements in set S

H(t) - Entropy of subset t

In ID3, information gain can be calculated (instead of entropy) for each remaining attribute. The attribute with the largest information gain is used to split the set S on this iteration.

6. EXPERIMENTAL SETUP AND RESULTS

To produce results for framework Weka 3.6 has been used.

The results have been compared after applying both concepts on following evaluation parameters.

a) Time taken to build model
b) Correctly classified instances
c) Kappa statistic
d) Mean absolute error
e) Root mean squared error
f) Relative absolute error
g) Root relative squared error
h) Detailed Accuracy By Class
i) Confusion Matrix

Table 1 Evaluation on different parameters of both approaches

S.no.	Evaluation Parameters	After Numeric to binary filter	After Discretization
1.	Time taken to build model	1.15	0.92
2.	Correctly Classified Instances	98.9163 %	99.4324 %
3.	Kappa statistic	0.9793	0.9945
4.	Mean absolute error	0.0138	0.0027
5.	Root mean squared error	0.0919	0.0524
6.	Relative absolute error	2.7672 %	0.5535 %
7.	Root relative squared error	18.4199 %	10.5214 %

Table 2.Detailed accuracy by class.

Class	After Numeric to binary filter			After Discretization		
	TP Rate	FP Rate	Preci-sion	TP Rate	FP Rate	Preci-sion
normal	0.989	0.009	0.992	0.997	0.003	0.998
anomaly	0.991	0.011	0.987	0.997	0.003	0.998
Weighted Avg.	0.99	0.01	0.99	0.997	0.003	0.998

Table 3 Confusion Matrix

	After Numeric to binary filter		After Discretization	
Class	**Normal**	**Anomaly**	**Normal**	**Anomaly**
Normal	13291	153	11360	39
Anomaly	106	11628	30	11689

7. CONCLUSION & FUTURE WORK

By observing results discretization and classification with ID3, it's been concluded that in the proposed framework the discretization of data gives higher accuracy and fewer errors with less time consumptions. Which favors, this approach is more efficient for the Intrusion detection rather than directly application of classified algorithm or any other filtration like numeric to binary system. In future it may be enhanced by applying other classification algorithm or filters to preprocess the data.

8. REFERENCES

[1] Litty Lionel, "Hypervisor-based Intrusion Detectio", Master of Science Graduate department of computer Science University of Torronto, 2005.

[2] http://www.mendeley.com/research/naive-bayes-vs-decisiontrees-in-intrusion-detection-systems/#page-1

[3] SrinivasMukkamalaa, Andrew H. Sunga and AjithAbrahamb Intrusion detection using an ensemble of intelligent paradigms‖; www.elsevier.com/locate/jnca, January 2004.

[4] "Data Mining Concepts and Techniques" byJiawei Han and MichelineKamber from Morgan Kaufman Publications.

[5] Adriaan; "Introduction to Data Mining",Addison Wesley Publication

[6] A.K.Pujari; "Data Mining Techniques"; University Press

[7] Salem,karim "Revising the outputs of a decision tree with expert knowledge: Application to intrusion detection and alert correlation", 2012, ieee, p 452 -459.

[8] Anupama Mishra, B. B. Gupta, R. C. Joshi," A Comparative study of Distributed Denial of Service Attacks, Intrusion Tolerance and mitigation Techniques" European Intelligence and Security Informatics Conference-2011.

[9] Saketh Kumar shakkariG.Varalakshmi, "Detection of application layer DDOS attack for a popular website using delay of transmission", INTERNATIONAL JOURNAL OF ADVANCED ENGINEERING SCIENCES AND TECHNOLOGIES Vol No. 10, Issue No. 2, 181 – 184-2011.

[10] GU Jun. "Research on intrusion detection system based on KPCA and SVM". Journal of Computer Simulation, 2010, 27(7): 105-107.

[11] Xiong Wen, Wang Cong. "Hybrid feature transformation based on modified particle swarm optimization and support vector machine". Journal of Beijing University of Posts and Telecommunications, 2009, 32(6): 24-28.

[12] LI Zhong-long, SI Jin. "Distributed Denial of Service Analysis". Computer Knowledge and Technology, 2010, 6 (6): 2373-2374.

[13] Xiang Xu ,Ding Wei , Yuelei Zhang. "Improved detection approach for DDOS attack based on SVM", 2011, IEEE

[14] P. K. Agrawal , B. B. Gupta , Satbir Jain . "SVM Based scheme for Predicting Number of Zombies in a DDoS Attack" 978-0-7695-4406-9/11 $26.00 © 2011 IEEE

[15] KashifSaghar , William Henderson ,David Kendall , Ahmed Bouridane , "Applying formal modeling to detect Dos attack in wireless medium"

[16] Xinfeng Ye , Santosh Singh " A soa approach to counter ddos attack" 2007 IEEE\

[17] http://en.wikipedia.org/wiki/ID3_algorithm

[18] U.M.Fayyad and K.B.Irani, "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning," Proc. 13th Int. Joint Conf. Artificial Intelligence, pp. 1022-1027, 1993.

[19] B.Hemada, K.S.Vijaya Lakshmi "Discretization Technique Using Maximum Frequent Values and Entropy Criterion" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2013.