International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Bayesian Networks for Intrusion Detection Systems using K2 algorithm

A. Anthony Paul Raj [1] ,J. K. Kani Mozhi [2]

.

[1] *Research Scholar (Part Time), Periyar University, Salem,(India)*

[2] *Professor, Dept.of Computer Applications, Sengunthar Arts and Science College, Salem Road,Tiruchengode, Namakkal,(India)*

ABSTRACT

Those expanding utilization of internet has significantly contributed to the rising wide variety of attacks that inhabit within it. In search for a superior of Network security is Intrusion Detection Systems (IDS). The IDS have risen to emerge as a subject matter of research and situation in order to battle these attacks. The intention of IDS is to identify unsafe behavior that goals a network and its uncooked material.

In order to get together the requirement of growing records on web, IDS with Data Mining came into limelight helping in effective detection outcomes. The huge consequences in IDS with Data Mining have been seen through making use of Bayesian Network and K2 algorithm. A Bayesian Network (BN) is recognized as graphical representation tool used to mock-up decision difficulty containing improbability. BN and K2 gaining knowledge of along with open attacking device is used right here to make a computerized self-learning intrusion detection system. Hence the paper is the find out the focus on K2-based IDS in identifying the intrusions

Keywords - Intrusion Detection System (IDS), Bayesian Network (BN), K2 algorithm.

I. INTRODUCTION

Today, a giant quantity of sensitive statistics is processed via computer networks; consequently it is increasingly more vital to make records systems, especially these used for vital functions in the navy and industrial sectors, resistant and tolerant to community intrusions. IDS (Intrusion Detection System) strategies are used to strength safety and make bigger resistance to inside and exterior attack [1].Intrusion detection is now not an effortless undertaking Because of the vastness of the network undertaking data and the need to usually update the IDS to be adapted to unknown assault methods. Nowadays, absolutely defend a network from attacks is being a very tough task. Even closely protected networks are each from time to time penetrated, and Intrusion Detection. System looks to be crucial and is a solution portion in system and network security. IDS for information mining fundamentally based frameworks need the gain of probably being capable to observe new attacks and prevent

the assault on network. Data Mining performs a prominent role in the statistics analysis through extracting useful data from ever growing giant quantity of data. Data Mining, in addition widely submitted to as Knowledge Discovery from Data (KDD), is the programmed or useful mining of model representing information unreservedly saved or imprison in massive databases, information warehouses, the Web, different massive data repositories, or facts streams [2]. The located understanding is used by means of the IDS for detecting uncommon patterns of pastime that are acknowledged to correlate with intrusions. In continuance to beginning the paper wrap up the following sections: Section II give details about what are IDS and its types? Section III thrashes about Bayesian Network and K2 algorithm. Section IV displays the execution of IDS with K2 studying algorithm, Section and lastly Section V comprises conclusion and future extent.

II. INTRUSION DETECTION SYSTEM (IDS)

Intrusion detection System (IDS) is a type of protection software designed to robotically alert directors when anybody or something is challenging to negotiation data machine thru malicious things to do or through security policy violations. Cisco 2017 Annual Security Report [3]. Server-side vulnerabilities have been on the increase: Adversaries have realized that by use vulnerabilities in server software, they can gain larger get admission to organization networks. In the to start with a few months of 2017, server-side vulnerabilities show up to be on tune to show an increase of 36 percent from the quantity of 2016 vulnerabilities; client-side vulnerabilities show a possibly expand of 35 percent from 2016 (see Figure 1).

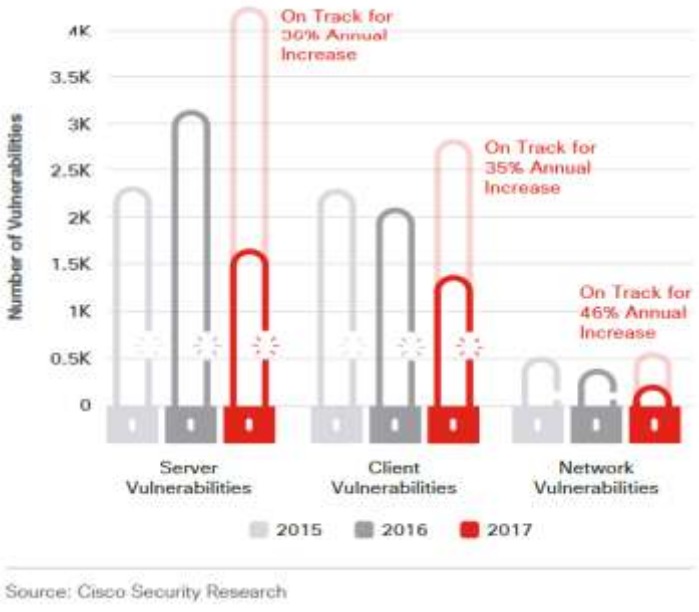

Figure1: Client – server vulnerabilities [3]

An IDS works through check system activity thru analyzing vulnerabilities in the system, the integrity of archives and conducting an analysis of patterns mainly focus on already regarded attacks. It also routinely

monitors the Internet to search for any of the modern-day threats which ought to end result in a future attack. Recognition Methods: Intrusion Detection System can only become aware of an attack. It can't stop the attacks. In contrast, IPS avoid assaults by using detecting them and stopping them earlier than they attain the target. An attack is any strive to compromise confidentiality, integrity, or availability.

The two important techniques of detection are signature-based and anomaly-based. Any kind of IDS (HIDS or NIDS) can discover assaults based totally on signatures, anomalies, or both. The HIDS video display units the network traffic attaining its NIC, and the NIDS video display units the site visitors to the network

1.1 Host-based intrusion detection system (HIDS)

A host-based intrusion detection system (HIDS) is extra software established on a machine such as a scheme or a server. It gives protection to the individual host and can discover conceivable assaults and defend critical operating gadget files. The fundamental intention of any IDS is to display traffic. HIDS is a host stand for innovative techniques in which a system bring together the data as the records of a mixture of actions of host with NT affair logs, system logs etc [4]. The predominant goal of any IDS is to display traffic. For a HIDS, this visitors passes through the community interface card (NIC).Many host support IDSs have developed to system application responsibility on the computer.

1.2 Network-based intrusion detection system (NIDS)

Network-based Intrusion detection system shows out the units' action on the networks, an administrator fit the NIDSs sensors on the public devices like routers and firewalls. These sensors build up the data and record to an essential screening server web hosting a NIDS console. A NIDS may be not capable to notice anomalies on person structures or workstations until the anomaly causes an important difference in network traffic. And in addition NIDS will be not skilled should unscramble what's more decrypt and encrypted movement. In different words, it can solely screen and check threats on the community from traffic sent in plaintext or non encrypted traffic.

III. BAYESIAN NETWORK

A Bayesian network is a graphic illustration of the common option for sharing function over a group of variables. The network formation is set for as a Directed Acyclic Graph (DAG) in which every node are related to an random variable and each part suggest a in need association between linked variables. Every variable (node) in a BN is linked with a Conditional Probability Table (CPT), which details the qualified possible for this variable given all the combos of its root node' values [2].

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

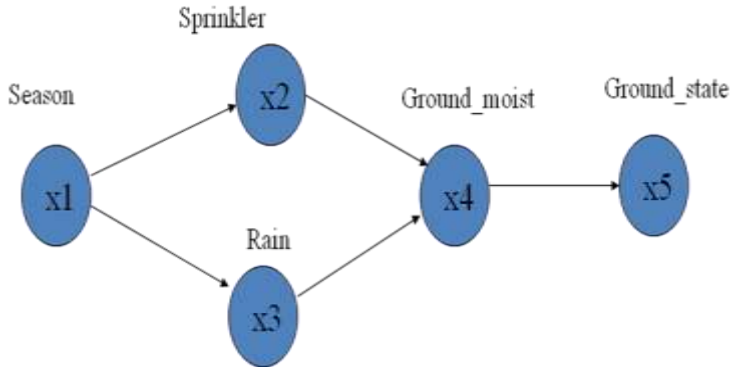

Figure 2: Bayesian Networks

Bayesian Networks [6] are defined as graphical probabilistic fashions for multivariate analysis. Specifically, they are directed acyclic graphs that have a related chance distribution characteristic [7]. Nodes in the directed diagram signify trouble variables (they can be both a premise and a conclusion) and the edges characterize conditional dependencies between such variables.

Regarding to its usual formulation, given two activities A and B, the conditional likelihood P (A|B) that A happens if B occurs can be acquired if we be aware of the chance that A occurs, P (A), the chance that B occurs, P (B), and the provisional probability of B given A, P (B|A) (as proven in equation 1). [8]

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Nodes inside the directed diagram symbolize trouble variables (they can be both a premise and a conclusion) and the edges symbolize conditional dependencies between such variables. Moreover, the probability characteristic illustrates the power of these associations in the layout [7]. (Figure 2)

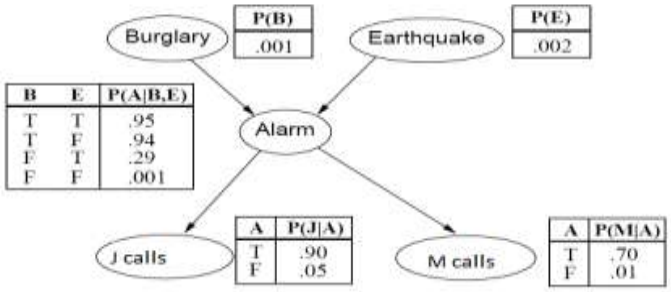

Figure 3: Example of a Bayesian Network

A Bayesian network organizing family members between occasions on the burglary-earthquake-alarm domain, are collectively with entire specs of all probability distributions.

IV. K2 ALGORITHM

K2 is a division of the regularly used algorithms underneath Bayesian Networks. It is an algorithm for developing BayesNet from a database of records. In seek algorithm that seek out for intrusions on the foundation of a build in ordered set of case. The algorithm requires a rest of nodes, a before recognized order on the nodes, a top certain on the wide diversity of root node may also have, and a database hold viable cases.
It starts with the aid of consider that a node has no roots, after which; in each step it provides increase the root whose accumulation in the main increases the likelihood of the ensuing structure [9]. The algorithm terminates adding mother and father to the nodes, when the accumulating of single root can't increment the probability of the network given the data.

Though the K2 algorithm works properly for intrusion detection along with exceptional error rate, it then again requires an ordered set described as heuristic and it additionally lacks in computational simplicity. The fundamental middle of thought of our algorithm is to attain a node ordering from data. This ordering can then be second-hand as an input parameter to the K2 algorithm, which will then learn the shape of the BN with higher accuracy.

Procedure for K2:

{ Input: A accept of n nodes, an ordering regarding the nodes, an upper bound u on the longevity variety regarding mother and father a node might also have, then a database D containing m cases.}

{Output: For each node, a printout of the root node}

1. Begin
2. i:=1;
3. $\pi_i := \theta$
4. Pold:= $g(x_i, \pi_i)$
5. J:= i+1;
6. Add (x_j, π_j)
7. Pnew =$g(x_j, \pi_j)$
8. If Pnew>Pold Then Pold: = Pnew
9. else Delete x_j, from π_j
10. j:= j+1;
11. If j>n Then i:= i+1;
12. Else go to step 3

13. If j>n; Then i = i+1;
14. Else go to step 5
15. If i== n Then End
16. Else go to Step 3

V. EXPERIMENTAL ANALYSIS OF K2

In each and every experiments are executed in WEKA reproduction tool by means of using KDDCup'99 dataset. Contains facts about proper and predicted classifications accomplished by means of a classification system. Performance of such structures is normally evaluated using the facts in the matrix. To consider our system, besides the classical accuracy measure, the three popular metrics of detection charge (DR) /True Positive Rate (TPR), false positive rate (FPR) & F-measure developed for community intrusions will be used.

Table 1 suggests these trendy metrics

		Predicted	
		Normal	**Attack**
Actual	Normal	**TP**	**FN**
	Attack	**FP**	**TN**

TABLE 1: CONFUSION MATRIX

Following are the class-wise overall presentation assessment of the K2 based IDS:

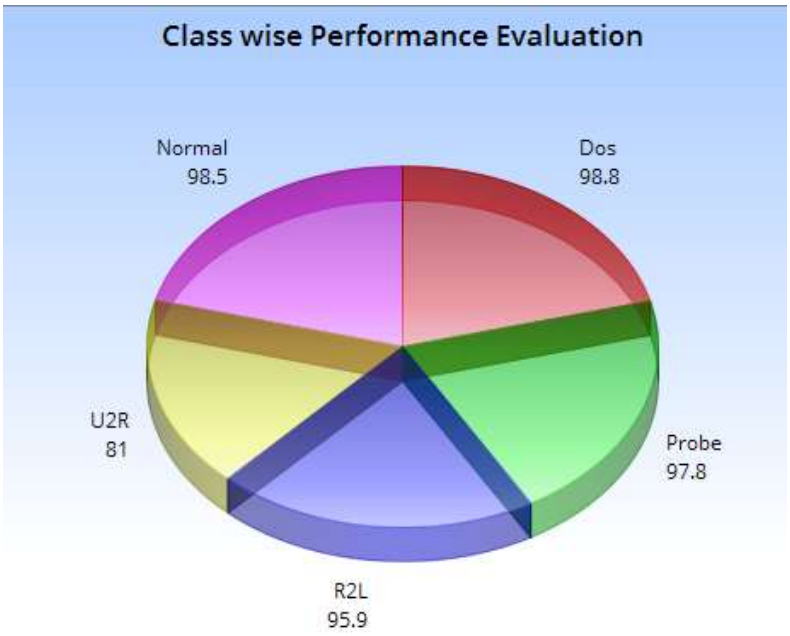

Figure 4: Class wise Performance Evaluation

VI.CONCLUSION

However the find out about indicates that the usage of facts removal with IDS comply with special methods on deck is well-organized and above this it established as a smart intrusion detection system technique. In addition to creating intrusion detection method greater fine along with limit false alarm rate, it is in a position to tackle IDS drawbacks such as classifying the assaults and correlating their types on the foundation of one-of-a-kind information records. Demand for improvisation therefore the need have been fulfilled to some expand by introducing fuzzy logic thought with the present systems. In contrast to this, there is additionally the extent for associated the statistics speculation such as Poisson distribution, Binomial distribution.

VI. ACKNOWLEDGEMENTS

Foremost, my gratitude goes to Dr. J. K. Kani Mozhi, Professor, Department of Computer Application, Sengunthar Arts & Science College, Tiruchengode, who guided me into the research work. Her vision, her enthusiasm and her spirit inspired and enlightened me to carry out this research paper with curiosity and involvement. And I thank my wife Mrs. A. Arockia Vinnarasi and my son Master A. Joel Roy and my daughter A. Jessica Mariam who always supported to me. Finally, I thank God who gives me good health, continuous encouragement and support throughout the research work and my life.

REFERENCES

[1] Jasreena Kaur Bains, Kiran Kumar, Kapil Sharma, Intrusion Detection System with Multi Layer using Bayesian Networks International Journal of Computer Applications (0975 – 8887),Volume 67– No.5, April 2013 No.

[2] Jiawei Han & Micheline Kamber, Data Mining:Concepts and Techniques, (2006), Second Edition, Morgan Kaufmann Publishers

[3] Cisco, Cisco 2017 Annual Security Report, 2017.

[4] Mradul Dhakar, Nisha Chaurasia, Akhilesh Tiwari , Analysis of K2 based Intrusion Detection System, Current Research in Engineering, Science and Technology (CREST) Journals.

[5] F. Jesen, Bayesian Networks and Decision Graphs, Springer, New York, USA, 2001.

[6] Pearl, J. & Russell, S. (2000). Bayesian networks, Technical Report Tech. Rep. R-216, ComputerScience Department, University of California, Los Angeles.

[7] Castillo, E., Gutiérrez, J. M. & Hadi, A. S. (1996). Expert Systems and Probabilistic Network Models, erste edn, Springer, New York, NY, USA.

[8] Abramson, B., Brown, J., Edwards, W., Murphy, A. & Winkler, R. L. (1996). Hailfinder: A Bayesian system for forecasting severe weather, International Journal of Forecasting: 57–71.

[9] Evelina Lamma and Fabrizio Riguzzi, Improving the K2 Algorithm Using Association Rule Parameters, Elsevier Publications (2005), pp. 1-11.