iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VIII **Month of publication:** August 2018

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VIII, August 2018- Available at www.ijraset.com

Sliding Window Based Malicious Traffic Detection in Network using Newcombs Law

Mr.Parmanand Sahu[1], Dr. Siddhartha Choubey[2]
[1, 2]*Shri Shankaracharya Group of Institution (SSTC), Dept. of Computer science & Engineering, Bhilai, Chhattisgarh, India*

Abstract: ***Network security assumes a key part in the lives of people and associations. Assailants turn to new and advanced strategies for executing assaults like DDos and Botnets with vindictive goal. In this paper we explore network traffic information utilizing natural law to recognize malicious and non-malicious network traffic and, to inspect and think about the execution. Dataset are chosen, arranged and isolated into little pieces utilizing the sliding window procedure. Newcomb law is contender for the examination, where law is connected to the isolated pieces with a specific end goal to recognize malicious and typical network traffic. Newcomb law had the best detection range between normal and malicious networks.***
Keywords: Intrusion Detection System, Network, Malicious Packets, Data Security, Newcomb's Law.

I. INTRODUCTION

The intrusion detection techniques based on information mining [1, 2] and large can be categorized as one of two classes: abuse detection and inconsistency detection. In abuse detection, each example in an informational index is named as 'ordinary' or 'nosy' and a learning calculation is prepared over the marked information. Research in abuse detection has concentrated predominantly on distinguishing network intrusions utilizing different order calculations [1], affiliation rules [1] and cost delicate displaying. Not at all like mark based intrusion detection frameworks, models of abuse are made naturally, and can be more advanced and exact than physically made marks. Irregularity detection calculations assemble models of ordinary conduct and consequently identify any deviation from it. The significant advantage of peculiarity detection calculations is their capacity to conceivably identify unexpected assaults. Also, they might have the capacity to recognize new or irregular, yet non-nosy, network conduct that is important to a network chief.

II. ANOMALY BASED DETECTION

The anomaly construct recognition is based with respect to characterizing the system conduct. The system conduct is as per the predefined conduct, at that point it is acknowledged or else it triggers the occasion in the anomaly location. The acknowledged system conduct is arranged or learned by the determinations of the system directors. The essential stage in characterizing the system conduct is the IDS motor ability to slice through the different conventions at all levels. The Engine must have the capacity to process the conventions and comprehend its objective. In spite of the fact that this convention investigation is computationally costly, the advantages it creates like expanding the manage set aides in less false positive alerts.

The major disadvantage of anomaly based detection is that it requires collecting and learning of rules and after that implementing those rules to IDS.

III. CLASSIFICATION OF ANOMALY BASED DETECTION

There are various types of anomaly based intrusion detection. Some of them are described below.

1) Statistical Based
a) Markov Process
b) Operational
c) Multivariate
d) Time Series
2) Cognition Based
a) Finite State Machine
b) Expert Systems
3) Machine Learning
a) Bayesian Network

©IJRASET: All Rights are Reserved

b) Genetic Algorithm
c) Neural Network
d) Fuzzy Logic

A. Statistical Based Model

Statistical based model includes Markov process model, time series model etc. These model are depend upon some statistical historic data which are processed by some formulas. These formulas are standard and are very accurate while producing results.

B. Cognition Model

This model is also known as expert system model. It require knowledge base for processing the information. The huge knowledge base is required for framework to predict the malicious flow of the network.

C. Machine Learning

This model is very popular among researches. There are lots of machine learning techniques are present which can be used for detection of malicious flow in network. The technique includes Bayesian Network simulation, neural network, fuzzy logic etc.
Table I. shows the comparison of various techniques used in various layers of network for malicious network detection.

TABLE I. Protocol and its Overhead in Various Layers for Malicious Network Detection System

Layer	Protocol	Overhead
Physical	RSSI Value	Calibration RSSI value for neighbor nodes
MAC	TDMA	Keeping track of TDMA for other nodes
Network	Checks neighbor and expected packet info matches	Updating of hops in the packets
Application	RTT	Calibration of RTT for each node

IV. LITERATURE SURVEY

Yi S et al. [7], presents the algorithm for information mining of intrusion discovery framework has been enhanced and streamlined to accomplish smart location of system information. Winsock2 SPI is utilized amid the outline to catch information in the system, and the strategy for "session sifting" is embraced to channel arrange bundles. The framework comprises of modules of control rules and wise discovery, and so on. As per real location, the framework is equipped for showing system association status consistently, adequately controlling application programs and shrewdly recognizing system information.
Jiankun Hu and Xinghuo Yu et al [8] attempts to improve the host-based anomaly intrusion identification, concentrating on framework call based HMM. This was again later improved with the consideration of information pre-handling for perceiving and dispensing with excess sub-groupings of framework calls, coming about in less number of HMM sub models. Exploratory outcomes on three open databases showed that preparation cost can be decreased by half without influencing the intrusion identification execution. False caution rate is higher yet sensible contrasted with the group preparing technique with a 58% information decrease.
R. Nakkeeran et al [9] proposed an anomaly discovery framework involving identification modules for identifying anomalies in each layer. The anomaly identification aftereffect of the neighbor node(s) is taken by the present hub and its outcome thusly is sent to the neighbor node(s).Experimental comes about uncovered expanded location rate and diminished false caution positives, contrasted with different strategies.
Jiong Zhang et al [10] proposed another structure of unsupervised anomaly NIDS in light of the exception location system in random forests calculation. The system manufactures the examples of system benefits over datasets marked by the administrations. With the fabricated examples, the system identifies attacks in the datasets utilizing the modified anomaly identification algorithm, decreasing the complexity. This approach is autonomous of free attack training datasets, however accept that each system benefit has its own particular example for ordinary activities.
TY. Zhao [11] develops a system intrusion identification display in light of information mining innovation, which can distinguish known intrusion adequately and has a decent ability to perceive obscure information diagram which can't be recognized successfully in customary IDS. The paper mostly does the accompanying work: by examining the intrusion profoundly, extricate the properties

©IJRASET: All Rights are Reserved

which can reflect intrusion attributes viably; consolidate abuse discovery, anomaly location and human mediation, build up rule library in view of C.45 decision tree calculation and utilize the ideal example coordinating in order to enhance identification rate.
Jabez Ja, et al. [12]. Proposes intrusion detection framework to recognize the attacks productively. Besides, it is similarly vital to identify attacks at a starting stage with a specific end goal to diminish their effects. This exploration work proposed another approach called exception identification where, the anomaly dataset is estimated by the Neighborhood Outlier Factor (NOF). Here, prepared model comprises of huge datasets with dispersed capacity condition for enhancing the execution of Intrusion Detection framework. The test comes about demonstrated that the proposed approach recognizes the anomalies adequately than some other methodologies.

V. METHODOLOGY

In this section we present proposed methodology in detail. Fig. 1. Shows the pictorial representation of complete process of malicious traffic detection.

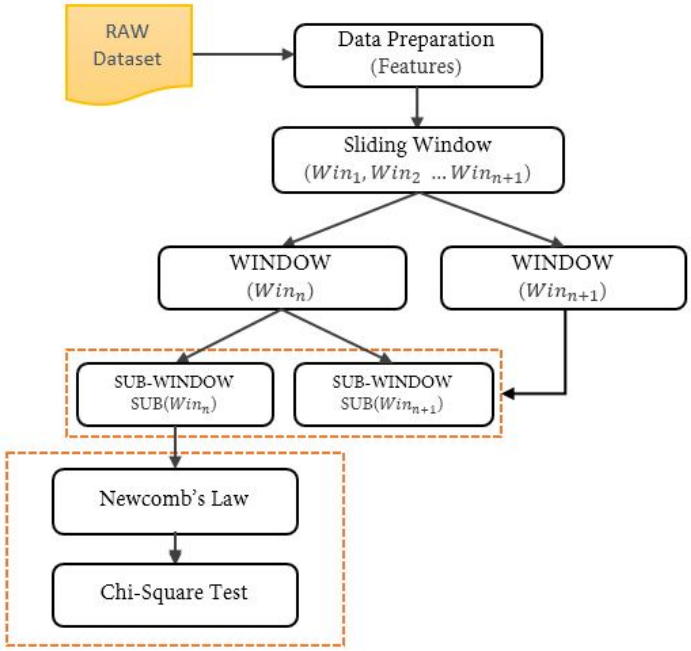


Fig. 1. Proposed System Architecture

A. RAW Dataset

The dataset which are used is KDD Dataset. It is build especially for detection of malicious traffic in the network.

B. Feature Extraction

From the dataset, there are a lot of irrelevant features are present which need to be removed. The feature represents the state of any system, hence selecting wrong or irrelevant feature may cause classifier to predict wrong results.

C. Sliding Window

After extraction features, these features are given as input to the sliding window phase. This phase simply arrange the dataset or features into various fixed length data. These data then subdivided into more sub windows. Then the divided data are given to Newcomb's law for analysis.

D. Newcomb Law

Newcomb's law, law of a anomalous numbers, and first-digit law, is a perception about the recurrence dissemination of driving digits in some genuine arrangements of numerical information. The law expresses that in numerous normally happening accumulations of numbers, the main critical digit is probably going to be small. For instance, in sets that comply with the law, the number 1 shows up as the most noteworthy digit around 30% of the time, while 9 shows up as the most huge digit under 5% of the

©IJRASET: All Rights are Reserved

time. By differentiate, if the digits were disseminated consistently, they would each happen around 11.1% of the time. Newcomb's law additionally makes expectations about the circulation of second digits, third digits, digit combinations, and so on.

A sliding window features is said to satisfy Newcomb's law if the leading windows Wd (d ∈ number of windows) occurs with probability

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(\frac{d+1}{d}\right) = \log_{10}\left(1+\frac{1}{d}\right)$$

P(d) is the probability of the window – which contains the features are malicious or non-malicious.

VI. RESULT

In this section experimental results are presented. The dataset used are KDD dataset. The snapshot of the dataset is shown in fig. 2.

```
0,tcp,http,SF,181,5450,0,0,0,0,0,1,0,0,0,0,0,0,0,0
,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,9,
9,1.00,0.00,0.11,0.00,0.00,0.00,0.00,0.00,normal
.
0,tcp,http,SF,239,486,0,0,0,0,0,1,0,0,0,0,0,0,0,0,
0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,19,
19,1.00,0.00,0.05,0.00,0.00,0.00,0.00,0.00,norma
l.
0,tcp,http,SF,235,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0
,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,29
,29,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,norm
al.
```

Fig. 2. Shows the KDD Dataset Snapshot

The outcome of Newcomb law experimental results are shown in fig. 3.

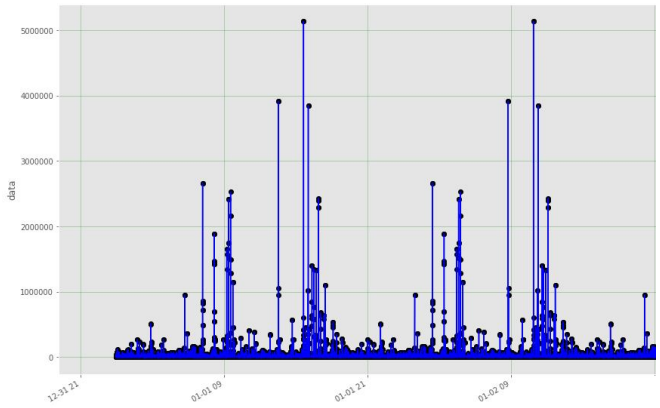


Fig. 3. Newcombs Law Output

In the above fig. X-Axis reprents the time series and Y-Axis represents the data. The high spark in the above graph represents the malicious traffic in the network.

VII. CONCLUSION

The rapid growth of networks has made open doors for intruders, (aggressors) who fall back on different abilities and procedures to execute network assaults. Network attacks like DDOS have been utilized as a part of different approaches to frustrate network administrations.

Network interruption recognition frameworks are not ready to distinguish attackers of this nature. In this paper we have demonstrated the Newcomb's law along with the Chi-Square test to predict the malicious traffic efficiently. Fig. 3, with contains the high spikes in the graph are the malicious traffic in that particular time.

REFERENCES

[1] Hazem M. El-Bakry, Nikos MastorakisA, "Real-Time Intrusion Detection Algorithm for Network Security,WSEAS Transactions on communications, Issue 12, Volume 7, December 2008

[2] Debar.H, Dacier.M and Wespi.A, "A Revised Taxonomy of Intrusion-Detection Systems" Annales des Telecommunications 55(7–8) (2000) 361–378

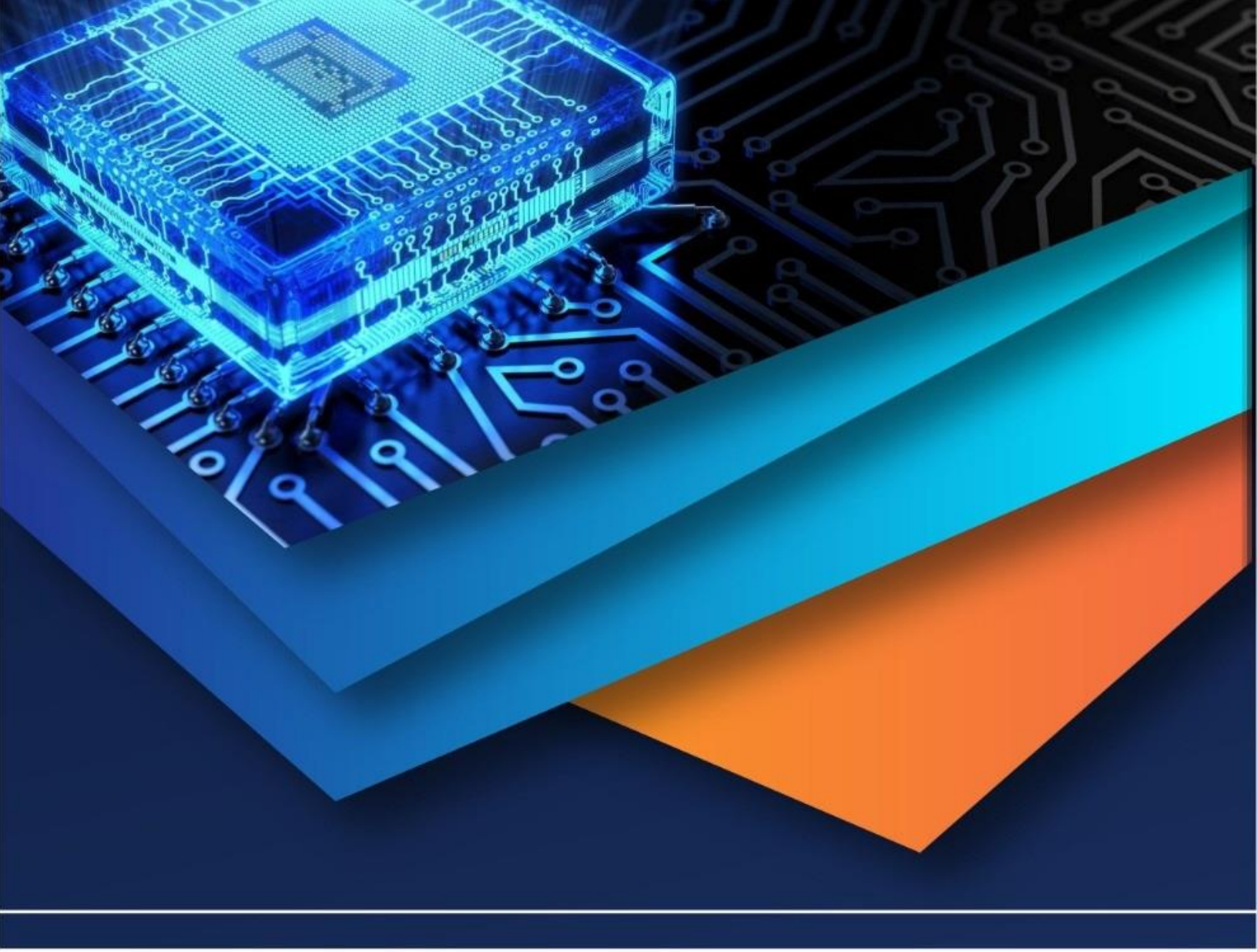
©IJRASET: All Rights are Reserved

[3] Roesch.M, “Snort - Lightweight Intrusion Detection for Networks” 13th USENIX Conference on System Administration, USENIX Association (1999) 229–238
[4] Allen.J, Christie.A, Fithen.W, McHugh.J, Pickel.J, Stoner.E, “State of the practice of intrusion detection technologies” Technical Report CMU/SEI-99TR- 028, Carnegie-Mellon University - Software Engineering Institute (2000).
[5] Hossein M. Shirazi,”Anomaly Intrusion Detection System Using Information Theory, K-NN and KMC Algorithms”, Australian Journal of Basic and Applied Sciences, 3(3): 2581-2597, 2009
[6] Wang. K and Stolfo.S.J, “Anomalous Payload-Based Network Intrusion Detection” 7th Symposium on Recent Advances in Intrusion Detection, Volume 3224 of LNCS., Springer-Verlag (2004) 203–222
[7] Yi S., Deng F. (2012) Research of Network Intrusion-Detection System Based on Data Mining. In: Gaol F. (eds) Recent Progress in Data Engineering and Internet Technology. Lecture Notes in Electrical Engineering, vol 157. Springer, Berlin, Heidelberg.
[8] Jiankun Hu and Xinghuo Yu, “A Simple and Efficient Hidden Markov Model Scheme for Host-Based Anomaly Intrusion Detection” IEEE Network Journal, Volume 23 Issue 1, January/February 2009.
[9] R. Nakkeeran, T. Aruldoss Albert and R.Ezumalai, “Agent Based Efficient Anomaly Intrusion Detection System in Ad-hoc networks” IACSIT International Journal of Engineering and Technology Vol. 2, No.1, February, 2010.
[10] Jiong Zhang and Mohammad Zulkernine, “Anomaly Based Network Intrusion Detection with Unsupervised Outlier Detection” IEEE International Conference on Communications, 2006.
[11] Y. Zhao, "Network intrusion detection system model based on data mining," 2016 17th IEEE/ACIS International Conference on Software Engineering,Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Shanghai, 2016, pp. 155-160.
[12] Jabez Ja, Dr B.Muthu Kumar, "Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach", International Conference on Intelligent Computing, Communication & Convergence (ICCC-2015

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VIII, August 2018- Available at www.ijraset.com

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)