Journal of Theoretical and Applied Information Technology
31st December 2017. Vol.95. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ACCEPTANCE SAMPLING FOR NETWORK INTRUSION DETECTION

[1]C. MADHUSUDHANARAO, [2]Dr. M. M. NAIDU

[1]Research Scholar, Department of Computer Science and Engineering, SriVenkateswaraUniversity (SVU) College of Engineering, Tirupati -517502, India

[2] Professor and Dean, School of Computing, Veltech Dr.RR & Dr.SR University, Avadi, Chennai- 600062, India

E-mail: [1]masura_c@yahoo.com, [2]mmnaidu@yahoo.com

ABSTRACT

Network Intrusion Detection System (NIDS) is to prevent entry of anomalous network flows into networks. Hundred percent inspections of all the fragments of network flows for detecting malicious fragments and thereof anomalous flows are highly prohibitive. The Selective Sampling Method (SSM) considers only network flows of small size not exceeding 80 fragments. Further, it is applicable for detecting port scan and host scan attacks only. This study proposes a novel NIDS adapting acceptance sampling method, referred to as ASNID. It is applicable to detect Land, Xmass, Nestea, Rose, Winnuke, NULL Scan, Teardrop, Fraggle, Port scan, Host scan. A randomly chosen sample of fragments from a network flow is inspected for detecting whether it is anomalous or not. It reduces the computational effort by a factor of $0<k<1$ where k is the ratio of sample size to total fragments of a network flow. It is proved experimentally that the GMAI, performance metric of ASNID tends to one as the sample size increases to 60%. It is also proved that as the percentage of anomalous flows increases GMAI increases. Hence, ASNID would of immense use in network intrusion detection.

Keywords: *Acceptance Sampling, Selective Sampling, Geometric Mean Accuracy Index, Network Intrusion Detection, Network Attacks*

1. INTRODUCTION

The rapid growth of internet has increased cyber-crimes. Conventional mechanisms such as firewalls and authentication tools provide protective layer for secured networks. However, they are vulnerable to Denial of Service (DoS) and probe attacks. Network Intrusion Detection Systems (NIDS) augments for ensuring secured networks.

NIDS protects networks connected to the internet from malicious attacks by monitoring network flows predominantly at fragment level in network layer. Naïve methods proposed in the past inspect each fragment of a flow for detecting whether a fragment is not malicious by comparing with ominous fragment patterns. Obviously, the network flow is anomalous if at least one fragment is detected as malicious.

Due to large number of flows on a high-capacity links, inspecting all flows and each fragment is computationally prohibitive. The objective of the research work is to propose a method that maximizes accuracy and minimize computational effort. Androulidakis and Papavassiliou in [11] proposed Selective Sampling method (SSM) for Network Intrusion Detection to minimize the effort. In this method, flows of size not exceeding 80 fragments are selected for inspection to detect port scan and host scan attacks only.

To overcome the limitations of SSM, this study proposes a novel NIDS adapting acceptance sampling. The objectives considered in proposing the above method are to over DoS attacks viz. Land, Xmass, Nestea, Rose, Winnuke, NULL Scan, Teardrop, and Fraggle also in addition to port-scan and host-scan attacks and attain better accuracy in detecting anomalous flows irrespective of network flow size. The performance metric employed for evaluating the accuracy is Geometric Mean Accuracy Index (GMAI) [14].

It is found that the proposed method is more effective and efficient in detecting anomalous flows infected with wide range of malicious attacks irrespective of flow size. It is more effective than SSM because of its better GMAI. Further, it is efficient as it reduces significantly the computational effort compared to SSM. Hence, the proposed method deserves for consideration as a significant contribution.

The rest of the paper is organized as follows. Section 2 reviews the related work. The

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

problem statement is given in section 3. Section 4 presents the concepts of acceptance sampling. Section 5 proposes the adaption of acceptance sampling for Network Intrusion Detection. The performance evaluation of proposed method through simulation is presented in section 6. Section 7 presents analysis of experimental data. The contributions of the study and scope for future study are reported in section 8.

2. REVIEW OF RELATED WORK

The Transmission Control Protocol/Internet Protocol (TCP/IP) suite supports data transmission services for applications in computer networks. An intruder tries to gain unauthorized access to a network violating network security properties such as confidentiality, data integrity, availability and interrupt, intercept, modify or fabricate data.

The Network Intrusion Detection System is required for accurate detection of intrusion. An NIDS that use a set of rules for detecting known attacks is referred to as Misuse based NIDS (MNIDS).An NIDS that attempts to detect abnormal patterns in network flows is called as Anomaly based NIDS (ANIDS). An NIDS that combines the two approaches for detecting both known and unknown attacks is referred to as Hybrid NIDS (HNIDS).

Network intrusion detection based on naïve Bayesian classifier (NB)[1], support vector machine (SVM)[2] have been proposed by researcher
to classify the network flows as normal or anomaly. Yihua Liao et al [3] proposed k-nearest neighbor (KNN) classifier to classify TCP/IP sessions as normal or anomaly on DARPA BSD system call data.

Decision Trees (DT) such as C4.5[4], CART[6], soft computing techniques such as fuzzy logic model[8], Artificial Neural Networks (ANN)[5], genetic algorithm[7] and rule based methods such as snort [10] are also proposed by researchers.

Sandya Peddabachigari et al [4] presented hybrid model of C 4.5 and SVM to classify network intrusions on KDDCUP 99 data. The C 4.5 classifies the data in to DoS, Probe, R2L, U2R attack types and then SVM classify them into actual attack. The authors proved that the hybrid DT-SVM gives better performance than SVM.

Sheng Yi Jiang et al [15] investigated the detection accuracy of cluster based unsupervised intrusion detection (CBUID) method that uses nearest neighbor classification. A method that cascade k-Means clustering and the ID3 decision tree for classifying anomalous and normal packets is proposed in [16]. Similarly cascading k-means clustering and C4.5 [17], cascading k-means clustering and naïve Bayesian classifier [18] is proposed to improve the detection accuracy.

The performance of these methods is evaluated using bench marked datasets such as KDDCUP99, DARPA or real life data sets such as UNIBS, MAWI. The performance varies with respect to overall detection accuracy, detection accuracy for different attack types and high false positives.

The above network intrusion detection methods have the following limitations:

1) These methods inspect all fragments of a network flow to classify the flow as anomalous flow. The computational effort is high. When network load increases, number of fragments passing through the network also increases. This in turn increases computational effort.
2) The existing methods uses either network flow features such as source IP, source port, destination IP, destination port to detect network flow as anomalous or fragment features to detect the attacks at fragment level. The different layers of network are prone to various threats. Any compromise in a lower layer will affect all layers above it. Threats at fragment level and flow level are to be detected to achieve better performance.

The authors of [11] suggested that inspection of selective sample of flows is adequate to detect the anomalies using entropy, instead of inspecting all flows for network intrusion detection.

The novelty of this work is that it proposes the acceptance sampling to decide a network flow as anomalous by verifying a random sample of fragments of a flow. This work considered GMAI as performance metric as GMAI is most suitable metric for Network Intrusion Detection (NID) methods.

The task of designing a Network Intrusion Detection System for maximizing the accuracy of anomalous flows detection with minimum computational effort is still a challenging problem.

3. PROBLEM STATEMENT

A message for transmission from a source node to a destination node is transformed into a set of network flows in the transport layer of source node. Then, the network layer partitions each network flow into packets and in turn each packet into fragments. It is likely that an intruder transforms one or more fragments into malicious

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and also fabricates and inserts a few malicious fragments in the network flow.

A network flow containing malicious fragments is referred to as anomalous flow which damages network resources, interrupt the services or probes for knowing the state of nodes with ulterior motive. Generally, NIDS inspects each fragment of a network flow until a fragment is found malicious and classifies such network flow as anomalous flow. Otherwise, the network flow is classified as normal flow. Its computational effort is highly prohibitive, even though 100% classification accuracy is guaranteed.

Hence, it is perceived as a challenging problem to devise a method for detecting the anomalous flows with the conflicting objectives of maximizing the classification accuracy and minimizing the computational effort.

The scope of SSM [11] is limited to network flows of small size and detecting port scan and host scan attacks only. Further, its performance evaluation is not reported. Owing to this, it is felt worth to pursue for finding a new method for intrusion detection that shall cover other attacks in addition to port scan and host scan. Moreover, such a method shall minimize the computational effort and maximize the GMAI.

4. ACCEPTANCE SAMPLING

The quality of a product depends on the quality of incoming raw materials / parts and conversion process thereof. For preventing the entry of inferior quality raw materials / parts into a manufacturing system, employing 100 % inspection of a lot that necessitates destructive testing is absurd. Similarly, the cost of 100 % inspection of a lot employing non-destructive testing is highly prohibitive. Alternatively, acceptance sampling is employed to decide whether to accept a lot.

In acceptance sampling, a sample of size 'n' chosen randomly from lot of size 'N' is inspected. A lot is accepted provided the number of defectives are less than or equal to c, the acceptance number. The ideal as well as usual behavior of the probability of acceptance of a lot as a function of the proportion of defectives is depicted in Figure.1, known as Operating Characteristic (OC) curves. The proportion of defectives acceptable to a customer is known Acceptable Quality Level (AQL). Similarly, the proportion of defectives not acceptable to a customer is known as Lot Tolerance Percent Defective (LTPD). It is obvious that in acceptance sampling, the probability of rejecting a lot with the proportion of defectives not exceeding AQL is not zero. Similarly, the probability of acceptance of a lot with proportion of defectives exceeding LTPD is not zero. The first possibility is referred to as producer's risk (α) whereas the second possibility is referred to as consumer's risk (β). The producer's risk and consumer's risk are also known as Type-I and Type-II errors respectively. A sampling plan (n, c) is designed with the objective of minimizing α and β, given N, AQL and LTPD.

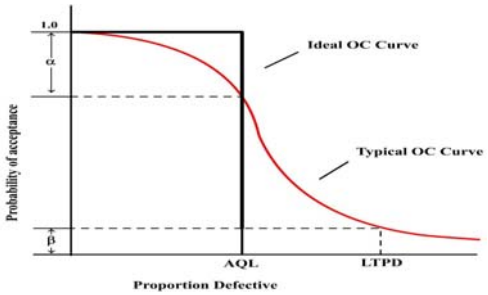

Figure 1 :Operating Characteristic Curve

The next section presents the existing Selective Sampling Method (SSM) for NID due to Androulidakis and Papavassiliou [11] and proposes acceptance sampling for NID.

5. ACCEPTANCE SAMPLING FOR NETWORK INTRUSION DETECTION

5.1 Selective Sampling Method:

The SSM selects a set of network flows for inspection employing flow size as threshold. The SSM is applied for detecting host scan and port scan attacks only. The port scan attack on a network flow is detected based on the entropy of flow computed using the Equation (1). Non-zero entropy indicates port scan attack.

$$E_p = -\sum_{i=1}^{n} P_i \log_2 P_i \qquad (1)$$

Where

E_p = Entropy of a network flow related to port scan attack.
x = number of fragments in the network flow
y = distinct number of destination ports
D = $\{d_i |$ distinct destination ports i where $1 \le i \le y\}$
P = $\{P_i$ is the probability of distinctive destination ports i, $1 \le i \le y\}$
U = $\{d_j |$ destination port j in a flow where $1 \le j \le x\}$

Similarly, the host scan attack on a network flow is detected based on the entropy of flow computed using the Equation (2). Non-zero entropy indicates host scan attack.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$E_h = -\sum_{i=1}^{n} P_i \log_2 P_i \qquad (2)$$

Where

E_h = Entropy of a network flow related to host scan attack.
x = number of fragments in the network flow
y = distinct number of destination hosts
D = $\{d_i |$ distinct destination hosts i where $1 \leq i \leq y\}$
P = $\{P_i$ is the probability of distinctive destination hosts i, $1 \leq i \leq y\}$
U = $\{d_j |$ destination host j in a flow where $1 \leq j \leq x\}$

The flow is classified as anomalous on detection of at least one of the above said attacks. Otherwise, it is classified as normal flow.

5.2 Acceptance Sampling:

It is obvious that 100% inspection of the fragments of network flow guarantees 100 % accuracy in detecting anomalous flows. However, the highly prohibitive computational effort affects application response time adversely. The selective sampling method is applicable to network flows of small size for detecting port scan and host scan attacks only.

Hence, it is perceived as worth to investigate the applicability of acceptance sampling for network intrusion detection. The parameters of acceptance sampling in statistical quality control and network intrusion detection are given in Table 1.

Table 1: Parameters of Acceptance Sampling in Statistical Quality Control and Network Intrusion Detection

Parameter	Notation	Statistical Quality Control	Network Intrusion Detection
Lot size	N	Number of units in a lot submitted for inspection	Number of fragments in a network flow
sample size	n	Number of randomly chosen units for inspection	Number of randomly chosen fragments for inspection
acceptance number	c	Threshold of defective units in a sample for accepting a lot, usually nonzero	Threshold of malicious fragments in a sample ought to be zero for a normal flow. Otherwise, the flow is detected is anomalous
% defective / anomalous	p	Percent defectives in a lot submitted for inspection	Percent malicious fragments in a network flow submitted for inspection

The steps of overall logic of acceptance sampling for detecting anomalous flows are as follows:

1. Initially assume that the network flow under consideration as a normal flow
2. Choose a random sample of fragments of size 'n' randomly from the network flow
3. Inspect the first / next fragment
4. If the fragment is detected as malicious then classify the flow as anomalous and stop
5. Otherwise go to step 3 while the next fragment exists

6. PERFORMANCE EVALUATION

This section presents performance metric, performance analysis and performance measurement of acceptance sampling for network intrusion detection. Further, its performance is compared with that of selective sampling method for detecting port scan and host scan attacks only.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6.1 Performance Metric

The Geometric Mean Accuracy Index proposed by C. Madhusudhanarao and M.M.Naidu [14], is employed as metric for evaluating the performance of Acceptance Sampling Network Intrusion Detection (ASNID) method. GMAI is computed using the equation (3), the geometric mean of True Positive Rate (TPR) and True Negative Rate (TNR).

$$GMAI = \sqrt{TPR \times TNR} \qquad (3)$$

Where TPR is the proportion of anomalous flows correctly detected and TNR is the proportion of Normal flows correctly detected.

6.2 Performance Analysis

The 100% Inspection as well as ASNID method search linearly for detecting malicious fragments. Hence, their asymptotic time complexities are as shown in Table 2.

Table 2: Asymptotic Time Complexities

Method	Best Case	Worst Case	Average Case
100 % Inspection	O(1)	O(N)	O(N)
SS	O(1)	O(N)	O(N)
ASNID	O(1)	O(n)	O(n)

It is obvious that the computational effort of ASNID method is $k \times O(N)$ where $0 < k < 1$. Further, the SSM is applicable to network flows of small size for detecting host scan and port scan only.

6.3 Performance Measurement

A simulation model for performance measurement of ASNID is developed. The synthetic network flow data set generated using a model proposed in [14], is applied. The details of experimentation using simulation model is presented here under.

The synthetic network flow database comprises three relations given in below:

(i) NFS is a relation of synthetic network flow. Its definition follows:

NFS = {<FLN, NoF, FCL> | FLN, NoF and FCL are flow number, number of fragments and flow class label respectively}

The primary key is {FLN}.

n (NFS) =Cardinality of a synthetic network flow set.

n (NFS) is constant and fixed as 10,000. However, the number of fragments in a network flow depends on network flow size, which varies. The flow class label is binary, normal or anomalous. Further, the percentage of anomalous flows in a network flow set is varied between 10% and 90% in steps of 10%.

(ii) FR is a relation, which provides fragment patterns of NFS. It definition follows:

FR = {<FLN, FRN, FRP> | FLN, FRN and FRP are flow number, fragment number and fragment pattern respectively}

The primary key is {FLN, FRN}. The foreign key is FLN that references NFS.

In case of an anomalous flow, the malicious fragments are randomly distributed and the percentage of malicious fragments is constant and fixed as 20%.

(iii) AP is another relation, which provides attack pattern data. Its definition follows:

AP = {< nc, ty, ap> | nc, ty and ap are nomenclature, type and pattern of an attack respectively}
m (AP) = cardinality of set AP

The domain of attribute **nc** is (Land, Xmass, Nestea, Rose, Winnuke, NULL Scan, Teardrop, Fraggle, Port scan, Host scan). The domain of attribute **ty** is (Denial of Service (DoS), Probe).

The input factors are sample size and percentage of anomalous flows of a synthetic network flow set. The sample is given as a percentage of fragments in a network flow. The number of equally spaced levels of sample size factor is 10 in the range of [10%-100%]. The number of equally spaced levels of percentage of anomalous flows factor is nine in the range of [10%-90%]. It necessitates conducting 90 experiments.

For each synthetic network flow set of cardinality 10,000, simulation is performed to produce confusion matrix.

FSS = {fi | fi FRP ($1 \le i \le s$ (FSS)}

Where,
FSS = fragment sample set chosen

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

randomly from a network flow
S(FSS) = cardinality of sample
SFCL = $\begin{cases} "Normal" if\ f_i not \in AP \forall i \\ "Anomalous" Otherwise \end{cases}$

Where,
SFCL = Flow class label due to simulation

The simulations are performed taking a synthetic network flow set consisting of a specific percentage of anomalous flows as input and varying the sample size in the range of [10%-100%]. The elements of confusion matrix shown in Table 3 are incremented by 1 as per rules given in Table 4. Initially, the elements of confusion matrix are set to 0 for a given combination levels of the percentage of anomalous flow and sample size. The pseudocode of ASNID method is given in Appendix. I.

Table 3: Confusion Matrix

		Simulation Flow Class Label	
		Anomalous	Normal
Actual	Anomalous	TP++	FN++
Flow Class Label	Normal	FP++	TN++

Table 4: Rules for Increment of Elements of Confusion Matrix

Rule No.	Rule Proposition
1	In case of anomalous flow, increment the count of True Positive (TP) by 1 if SFCL does conform.
2	In case of anomalous flow, increment the count of False Negative (FN) by 1 if SFCL does not conform.
3	In case of normal flow, increment the count of False Positive (FP) by 1 if SFCL does not conform.
4	In case of normal flow, increment the count of True Negative (TN) by 1 if SFCL does conform.

Ninety experiments are conducted, GMAI is computed for each experiment and the outcome of the experiments is shown in the Table 5. Its graphical representation is shown in Figure 2 and Figure 3. The analysis of experimental data is presented in the next section.

Table 5: Sample Size Vs Percentage of Anomalous Flows

% Anomalous Flows	10	20	30	40	50	60	70	80	90
sample size in %									
10	0.998008	0.998105	0.998286	0.998422	0.998781	0.998927	0.998959	0.998982	0.999324
20	0.998989	0.999423	0.999598	0.999717	0.99979	0.999821	0.999874	0.999881	0.999888
30	0.999666	0.999751	0.999785	0.999856	0.999875	0.999888	0.999904	0.999915	0.999938
40	0.999834	0.999901	0.999904	0.999915	0.999928	0.999929	0.999938	0.999945	0.999962
50	0.999937	1	1	1	1	1	1	1	1
60	1	1	1	1	1	1	1	1	1
70	1	1	1	1	1	1	1	1	1
80	1	1	1	1	1	1	1	1	1
90	1	1	1	1	1	1	1	1	1
100	1	1	1	1	1	1	1	1	1

Journal of Theoretical and Applied Information Technology
31st December 2017. Vol.95. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

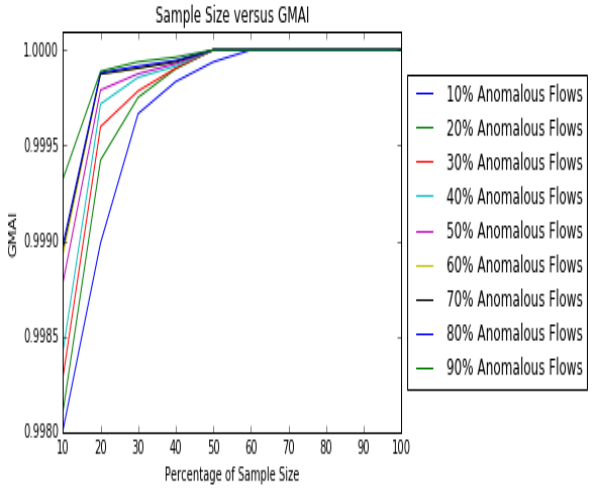

Figure 2: Percentage of Sample Size Vs GMAI

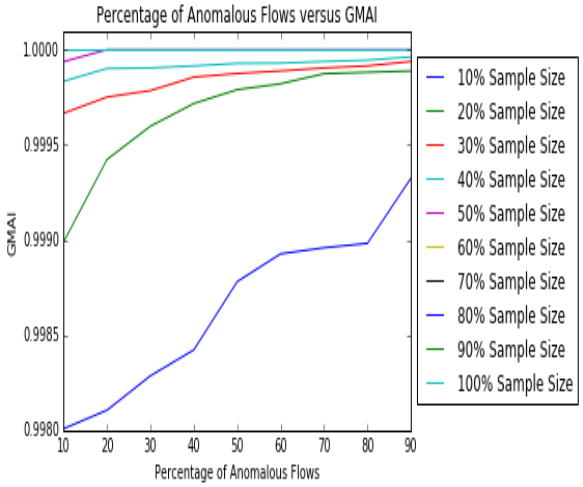

Figure 3: Percentage of Anomalous Flows Vs GMAI

7. ANALYSIS OF EXPERIMENTAL DATA

It is evident from Figure 2 that for a given network flow associated with a specific percentage of anomalous flows, as the percentage of sample size increases, the GMAI monotonically increases to one. From Figure 3, it is obvious that for a given sample size, as the percentage of anomalous flows of a set of network flows increases the GMAI monotonically increases to one.

Similarly for a given sample size, the GMAI monotonically increases to one as the percentage of anomalous flows of network flows increases. Likewise, for a given percentage of anomalous flows, the GMAI monotonically increases to one as sample size increases.

The above findings are significant in making the decisions given hereunder:

a) The percentage of anomalous flows of a set of network flows is an empirical probability that a network flow is anomalous, referred to as anomalous probability. Given the anomalous probability and service level represented by GMAI, facilitates to decide an appropriate sample size.
b) Similarly, given the sample size and anomalous probability, facilitates to predict GMAI.

The following also observed from the experimentation of ASNID method:

1) The Type-I error i.e., classification of normal network flow as anomalous is zero for all sample sizes of fragments.
2) The Type-II error i.e., classification of anomalous network flow as normal is decreasing as sample size increases and hence GMAI is increased.

Nine experiments have been conducted using simulation model for evaluating the performance of Selective Sampling Method for Network Intrusion Detection employing the same nine sets of synthetic network flows. The performance of ASNID method and SS method is tabulated in Table 6 depicting the same in Figure 4.

Table 6: Performance Comparison of SS Method and ASNID Method apropos GMAI

Fragment Sample size		Method			
		SSM	ASNID		
		100%	60%	50%	10%
Anomalous Probability (%)	10	0.996101	1	0.999937	0.998008
	20	0.997077	1	1	0.998105
	30	0.997106	1	1	0.998286
	40	0.997305	1	1	0.998422
	50	0.997328	1	1	0.998781
	60	0.997508	1	1	0.998927
	70	0.997973	1	1	0.998959
	80	0.997751	1	1	0.998982
	90	0.996955	1	1	0.999324

It is evident from Table 6 and Figure 4 that the GMAI of ASNID reaches to one for a fragment sample size of 60% drawn from the entire set of flows irrespective of the anomalous

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

probability. Similarly, for a sample size of 50% GMAI reaches to one except for anomalous probability of 0.1. It is found that the GMAI of ASNID is greater than that of SSM for a sample size of 10% for all levels of anomalous probability.

.

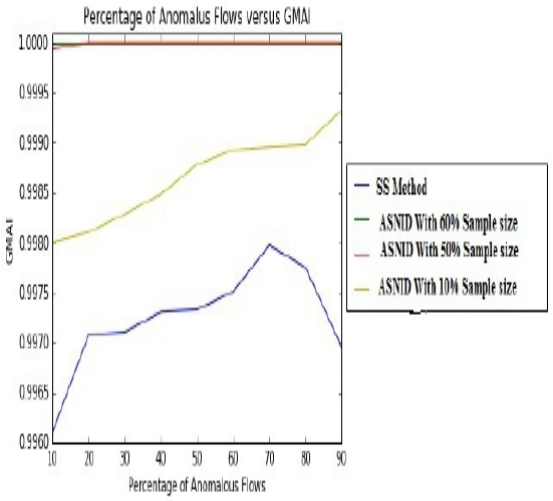

Figure 4: Performance Comparison of SSM Vs ASNID

If the same network flows are to be inspected, SS method inspects all fragments to classify a flow as anomalous or normal. However, ASNID inspects only a sample of fragments and classify the flow. The computational effort is reduced in ASNID method. Thus, the ASNID method maximizes the GMAI and minimizes the inspection effort.

The ASNID method is more effective and computationally efficient. It is capable of detecting land, nestea, rose, WinNuke, null scan, Xmass, fraggle, teardrop, port scan and host scan attacks.

The SS method is relatively less effective and computationally less efficient. It can detect port scan and host scan attacks only.

8. CONCLUSIONS

The objective of study is to propose a method ensuring it better than SSM with respect to GMAI and computational effort, capable of detecting broad range of attacks irrespective of network flow size. It is proved experimentally employing a synthetic dataset [14] that the proposed ASNID method is better than SSM as it yields better GMAI and reduces computational effort irrespective of network flow size. Further, it detects ten attacks viz.

The proposed method inspects fragments chosen randomly from a network flow employing acceptance sampling for detecting whether the flow is anomalous. The further study could be to investigate the utility of extending acceptance sampling for choosing network flows randomly. It can also be attempted to enlarge the scope coverage of attacks.

9. REFERENCES

[1] Christopher Kruegel, Darren Mutz, William Robertson ,Fredrik Valeur, " Bayesian Event Classification for Intrusion Detection" *Proceedings of 19th Annual Computer Security Applications Conference,* Las Vegas, USA, December, 8-12, 2003,pp.14-23.

[2] S. Mukkamala, G. I. Janoski, and A. H. Sung. "Intrusion Detection Using Support Vector Machines", *Proceedings of the High Performance Computing Symposium - HPC*, San Diego ,USA,April,2002, pp 178-183.

[3] Yihua Liao, V. Rao Vemuri, "Use of K-Nearest Neighbor classifier for Intrusion Detection", *Journal of Computers & Security*, Vol 21, No 5, 2002, pp 439-448.

[4] Sandya Peddabachigari, Ajith Abraham, Crina Grosan, Johnson Thomas, " Modeling Intrusion Detection System using Hybrid Intelligent Systems" *Journal of Network and Computer Applications,* Vol 30, No.1, 2007, pp 114-132.

[5] Srinivas Mukkamala, Andrew H. Sung "Feature Selection for Intrusion Detection using Neural Networks and Support Vector Machines" *Journal of transportation research board*, Vol 1822, 2003,pp1822-5.

[6] Srilatha Chebrolu, Ajith Abraham, Johnson P. Thomas, "Feature deduction and ensemble design of intrusion detection systems" *Journal of Computers & Security,* Vol 24, 2005,pp 295-307.

[7] Taeshik Shon**,** Yongdue Kim**,** Cheolwon Lee, Jon sub Moon, "A Machine Learning Framework for Network Anomaly Detection using SVM and GA" *Proceedings of the IEEE Workshop on Information Assurance and Security,* United States Military Academy, West Point, NY,June,15-19, 2005,pp 176-183.

[8] Surat Srinoy, Werasak Kurutach, Witcha Chimphlee, Siriporn Chimphlee, Santi Sounsri, "Computer Intrusion Detection with Clustering and Anomaly Detection Using ICA and Rough, Fuzzy" *Proceedings of the 4th WSEAS Int. Conference on Information Security,*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Communications and Computers, Tenerife, Spain, December, 16-18, 2005, pp 252-257.

[9] Source fire, Inc., Snort: The Open Source Network Intrusion Detection System. http://www.snort.org, 2007.

[10] M. Roesch, "Snort - Lightweight Intrusion Detection for Networks," *Proc. 13th USENIX Conference on System Administration,* Seattle, Washington, November, 7-12, 1999, pp 229–238.

[11] Georgios Androulidakis, Vassilis Chatzigiannakis, Symeon Papavassiliou, "Network Anomaly Detection and Classification via Opportunistic Sampling", *IEEE Networks,* Vol 23, No.9, 2009, pp 6-12.

[12] N. G. Duffield, P. Haffner, B. Krishnamurthy, and H. Ringberg, "Rule-Based Anomaly Detection on IP Flows," *Proc. 28th IEEE International Conference on Computer Communications,* Joint Conference of the IEEE Computer and Communications Societies. Rio de Janeiro, Brazil, April, 19-25, 2009, pp 424–432.

[13] R. E. Schapiro, "A brief introduction to boosting," *Proc. 16th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, 1999, pp1401–1406.

[14] C. Madhusudhanarao, M. M. Naidu , "A model for generating synthetic network flows and Accuracy index for evaluation of anomaly network intrusion detection systems*", Indian Journal of Science and Technology,* Vol 10 No.14, 2017,pp 1-16.

[15] S Jiang, X Song, H Wang, J J Han, Q H. Li, "A Clustering Based Method for Unsupervised Intrusion Detections," *Pattern Recognition Letters*, Vol. 27, no. 7, 2006, pp. 802–810.

[16] Shekhar R. Gaddam, Vir V. Phoha, Kiran S. Balagani, "K-Means+ID3: A Novel Method for Supervised Anomaly Detection by Cascading K-Means Clustering and ID3 Decision Tree Learning Methods", *IEEE Transactions on knowledge and data Engineering* , Vol.19,No.3, 2007, pp 1-10.

[17] Amuthan Prabakar Muniyandi, R. Rajeswari, R. Rajaram "Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree algorithm" *Procedia Engineering,* Vol.30, 2012, pp 174 – 182.

[18] Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir, "A K-means and naive bayes learning approach for better intrusion detection," *Information Technology Journal*., Vol. 10, No. 3, 2011, pp 648–655.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Appendix. I:

The pseudocode of ASNID is given below:

```
FLOW-CLASSIFICATION( )
1    TP← 0,FN←0,FP←0,TN←0
2    CM ← <TP,FN,FP,TN>
3    Read  k                          ▷  Percentage of fragments to be selected as sample
4    Read  p                          ▷  Percentage of anomalous flows
5    while (f(number-of-fragments ≠ eof))
       {
6        Read NFS                     ▷  read network flow record
7       <FLN, NoF, FL> ← NFS          ▷  FLN: flow number, NoF: Number of fragments, FL: flow label
8        Fcnt[FLN] ← NoF              ▷  array of fragment count
9        Flowlabel[FLN] ← FL          ▷  array of flow label
       }
10   while (synthetic-flow-data ≠ eof)
       {
11       if (flowno=Flsamp[i]         ▷  Array of flow numbers 1-10000
12          N ← Fcnt[flowno]          ▷  LOT size
13          n ← (N*k/100)             ▷  sample size
14          R← RANDSAMP( n,N)         ▷  array of random sample
15          pdp ←dp; pdip←dip
16          for  b=1 to N
17               Read fr
18               for c=1 to n
                   {
19                   if ( R[c]=b)
                       {
20                     frT ← PATTERNMATCH(fr)
21                     if(dp=pdp)
22                        dpc←dpc+1
23                     if(dip=pdip)
24                        dipc←dipc+1
                       }
25          if (dpc=n)
26             dpt ←1
27          if (dipc=n)
28               cpt ←1
29          if( dpt=1||cpt=1)
30               tt ←1
31          else
32               tt ←0
33          if( tt=1||frT=1)
34               ft ←1
35          else
36               ft ←0
37          CM← COMPUTE(fT, CM)
       }
       }
38   <TP,FN,FP,TN> ← CM
39   TPR ← TP/(TP+FN)
40   TNR ← TN/(TN+FP)
41   GMAI ← SQRT(TPR*TNR)
```

Figure 5: ASNID Method

The lines 1 and 2 in Figure5 are initialization statements. Line 3 reads percentage of fragments of a network flow to be selected as sample size. Line 4 reads Maximum number of flows. Lines 5-9 reads fragment count and flow label of each flow and stores in Fcnt, Flowlabel arrays. Lines 10-36 verify the flow type. N

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in Line 12 is number of fragments and it is considered as LOT. Line 13 gives the sample size. Line 14 randomly generates n fragment numbers between 1 to N and stores them in an array R. Lines 16-20 verifies rules of intrusion for each selected fragment by calling PATTERNMATCH function. Lines 21-24 increment dpc and dipc if destination port, destination host are same as before identify flow level attacks. Lines 25-36 compute a value for flow type (ft). If ft=1 flow is anomalous else normal. Lines 38-41 compute GMAI.

```
PATTERN-MATCH(fr)
1     if (Proto=1)
2       frT← TCP-SEARCH(fr )
3     else
4       frT← UDP-SEARCH ( fr )
5     Return frT
```

Figure 6: Pattern Matching

Lines 1-5 in Figure6 call TCP-SEARCH function if protocol is TCP else call UDP-SEARCH function and returns a value one for frT if fragment is malicious else zero.

```
TCP-SEARCH(fr)
1     k← FO/185
2     if (SIP=DIP&&A=1&&S=1)
3       m←1
4     if (U=1&&P=1&&F=1)
5       m←1
6     if (MF=1&&Length!=1500)
7       m←1
8     if (U=1&&DP=139)
9       m←1
10    if ((FO-k*185!=0)
11      m←1
12    if (U=0&&A=0&&P=0&&R=0&&S=0&&F=0)
13      m←1
14    else
15      m←0
16    Return m
```

Figure 7: Tcp Attacks Detection

In Figure 7 Line 2 inspects LAND attack. In LAND attack, the source IP is same as destination IP, ACK, SYN flags are set to 1. In Line 4 the URG, PSH, and FIN flags are verified for set for Xmas tree attack. Line 6 examines rose attack. MF and total length are 1 and less than MTU respectively for Rose attack. Line 8 checks for Winnuke attack which sets urgent flag to1 and destination port to 139.Line 10 look at for nestea attack, the fragment offset is not either 0 or in multiples of 185. Line 12 inspects for null scan attack in which all flags are set to zero.

```
UDP-SEARCH(fr)
1     k← FO/185
2     if (S=1&&DIP=255)
3       m←1
4     if((FO-k*185!=0)
5       m←1
6     else
7       m←0
8     Return m
```

Figure 8: Udp Attacks Detection

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Line 2 inspects fraggle attack and line 4 checks for teardrop attack in Figure 8.

```
COMPUTE (fT, CM)
1     <TP,FN,FP,TN> ← CM
2     if (FL=Anomalous && FT=1)
3           TP ← TP+1
4     else If (FL=Anomalous && FT=0)
5           FN ← FN+1
6     else If (FL=Norma1 && FT=0)
7           FP ← FP+1
8     else If (FL=Norma1 && FT=1)
9           TN ← TN+1
10    CM ← <TP,FN,FP,TN>
11    Return CM
```

Figure 9: Computation Of Confusion Matrix.

In Figure 9 Line 1 initializes TP, FN, FP, and TN. Lines 2-3 increment TP if Flow label and flow type are anomalous. Lines 4-5 increments FN if Flow label is Anomalous and flow type is normal. Lines 6-7 increments FP if Flow label is Normal and flow type is anomalous. Lines 8-9 increments FN if Flow label is Normal and flow type is normal.

```
RANDSAMP( n,N)
1     for  a= 1 to n
2          R[a] ← durv(1,N)          ▷ discrete uniform random variate
3     Return R
```

Figure 10: Selection Of Fragment Numbers Of A Network Flow.

Lines 1-3 in Figure 10 generate n number of discrete uniform random variates between 1 and N and stores in an array R.