Scientia Africana, Vol. 13 (No.2), December 2014. Pp317-327
 ISSN 1118 - 1931

A SUPPORT VECTOR MACHINE APPROACH TO THE DEVELOPMENT OF AN INTRUSION DETECTION SYSTEM

C. Ugwu and A. Olowoyeye

Department of Computer Science,
University of Port Harcourt,
Port Harcourt, Nigeria.

Chidiebere.ugwu@uniport.edu.ng, *bimboyeye@yahoo.com*

Received: 01-09-14
Accepted: 20-10-14

ABSTRACT

This paper demonstrated the use of support vector machine (SVM) model to develop an Instruction Detection System (IDE) which detects attacks by classification in wireless local area networks. The implementation was done on real time environment where network packets captured were subjected to SVM model for classification as either an attack or a normal data. The classifications were performed by separating the data into different clusters using a hyperplane. Waterfall model software development methodology was used to develop the intrusion detection system application and implementation was carried out with java programming language. The model predictive ability was evaluated by modelling an attack type and comparing the results with a standard benchmark, 99.78% and 98.89% were obtained for the detection of normal and attacks when tested with 5500 packets. The results suggest an improved detection efficiency and false alarm rate.

Key words: Support Vector Machine, Classification algorithm, Network Security, Network Packets

INTRODUCTION

Information which is the most valuable asset of any organization is unsafe, resources are stolen and business secrets are exposed. These are the pains the organisations face in computer networks, the problems suffice in higher dimension in a wireless environment due to its open nature. All the efforts made to curb intruders from attacking the wireless network, seems not to be providing sufficient and effective control measure in reducing the negative practices. Majority of current security system mostly concentrates on encryption, firewall and access control, this alone cannot provide the security needed in these networks. By definition, Intrusion is any set of actions that attempts to compromise the integrity, confidentiality, or availability of a resource (Heady R et al, 1990).

A detection made by an instrusion detection system is usually a misuse detection and anomaly detection. The first is simpler than the second because it uses rule based or signature comparison methods in classifying it observations (Scarfone et al, 2007). The first intrusion detection system designed, were targeted on mainframe computers, and all users were local to the system considered. This greatly simplified the intrusion detection task, as interaction from

outside was rare. The intrusion detection system analyzed the audited information provided by the mainframe, either locally or on a separate machine, and reports security suspicious events (Stephen and Haystack, 1988). As the focus of computing shifted from mainframe environments to distributed networks of workstations, several prototypes of intrusion detection systems were developed to accommodate network issues. Here the first step was to get host-based intrusion detection systems to communicate (Jagannathan, 1993). In a distributed environment, users hop from one machine to another, possibly changing identities during their moves, and launching their attacks on several systems. Therefore, the local intrusion detection system on the workstation has to exchange information with its peers (Steven et al, 1991). In a network based system, or network intrusion detection system (NIDS), the individual network packets flowing through a network are analyzed. The NIDS can detect malicious packets that are designed to be overlooked by a firewall's simplistic filtering rules. In a host based system, the IDS examine the activity on each individual computer or host (Newman and Robert, 2009). Hybrid approaches have also been developed that uses both network based and host based intrusion detection tools in a multi host environment (Steven et al, 1991). As a side effect, more specialized intrusion detection tools have emerged that monitor the most critical elements of an organization's presence on the internet. These products monitor firewalls web servers or routers looking for evidence of attacks in the highly specific context of these network elements (WebStalker Haystack Labs, 1997) (Nitin et al, 2008).

There are many articles that are presented in the literature that discusses the techniques for intrusion detection. Among them is Baker who proposed a new agent based approach for intrusion detection using rough set based classification technique. This technique generates rules from a large database and has mechanisms through rough sets to handle noise and uncertainty in data (Bakar et al, 2008). Tweedale et al, (2009) proposed a neural network based Multi-Agent Classifier System (MACS) using trust, negotiation, and communication reasoning model for intrusion detection. The main contribution of their work was in a trust measurement method based on cognition and rejection rates. Wang and Chiang (2008) proposed a cluster validity measure with outlier detection and cluster merging algorithms for proving a support vector clustering algorithm. This algorithm is capable of identifying the ideal cluster numbers with compact and smooth arbitrary shaped cluster contours for increasing robustness of outliers and noises.

Tweedale et al, 2009 proposed a decision tree based algorithm to construct a multi class intrusion detection system which was used to improve the training time, testing time, and accuracy of intrusion detection system. Significantly, recent work in the area of content analysis has been conducted by Stefano Sanero whose research involved analysing individual packet pay loads using unsupervised learning techniques (Savaresi and Sanero, 2004). Their results highlighted the usefulness of packet inspection techniques in an unsupervised context. Labib and Vemuri developed a real time system called network based detector using SOM which was useful at detecting denial of service attacks (Vemuri and Labib, 2002). Moore and Zuev made use of

Scientia Africana, Vol. 13 (No.2), December 2014. Pp317-327
© College of Natural and Applied Sciences, University of Port Harcourt, Printed in Nigeria *ISSN 1118 - 1931*

statistical header based information to classify network traffic between categories such as WWW, mail, bulk, games etc (Zuev, 2006). Barbara and co-workers developed a tool called ADAM (Audit Data Analysis and Mining) which uses data mining techniques to create association rules to detect anomalies in TCP connections (Jajodia et al, 2001)

This paper aimed at developing an improved IDS to reduce false alarm and to find an effective and intelligent means of detecting known and unknown attacks, by using a hybrid of misused and anomaly approaches. This is contrary to what others have done where single systems were used.

MATERIALS AND METHODS

The system consists of a graphical user interface for the input which provides medium for capturing data from a wireless network, specifically, **pcap** (**p**acket **cap**ture) was used to capture packet on the network. It collected packet from the network, due to the volume of the network packets, some features were extracted in order to reduce the dimension of the data captured. Using the change in frequency for normal packet and attack packet we trained our intrusion detection system with normal and attack packets. This will enable the support vector machine model to classify packets coming in as normal or attack packet. WINPCAP and JPCAP were two packages used in capturing packets and interfacing with the operating system and network interface card.

Fig 1 depicts the architecture of instruction detection system, the events section learns from what it has being trained; support vector machine was used for classification. The support vector machine algorithm checks packets sent on the network by comparing it with the data it has been trained with during the training process. The classification will categorise the packets into normal or an attack, any significant deviation from the normal activities of the system is flagged as an attack, but if the packet conforms to the normal system activities it is a normal packet.

Feature extraction is the basis for high performance in an intrusion detection system. If the features are improperly selected, the performance of detection models will be greatly affected. This paper seek to remove the false alarm rate associated with intrusion detection system, in order to achieve this, some features were selected from each network packet, ten features were selected, these were used for training the detection system, this is shown in table 1.

In order to increase the efficiency of our model, we used a hybrid method of anomaly and misuse; this allowed us to include attack signature, into our designs. Four attack types were used; flooding attack, routing disruption, packet dropping attack, denial of service attack.

Table 1: Ten Features used for classification on the captured packets

No	Feature Names	Feature Description
1	Source IP	This is the address of a device that send packets on a network
2	Destination IP	The address of a device attached to an IP network (TCP/IP network) that receives packets
3	Protocol	A network protocol defines rules and conventions for communication between network devices
4	Source Port	A number assigned to user sessions and server applications in an IP network. The source port is a next-available number assigned by TCP/IP to the user's machine
5	Destination Port	Destination ports may be "well-known ports" (0-1023) for the major internet applications, such as web and e-mail. For example, all port 80 packets (HTTP packets) are directed to and processed by a web server
6	Payload Size	The Actual data in a network packet excluding the header.
7	Packet Count	The packet counter counts how many data packets that is sent/received in a network.
8	TCP Packet	Transmission control protocol packet
9	ICMP Packet	The Internet Control Message Protocol (ICMP) is one of the main protocols, it is used by network devices, like routers, to send error messages
10	UDP Packet	User Datagram Protocol packet

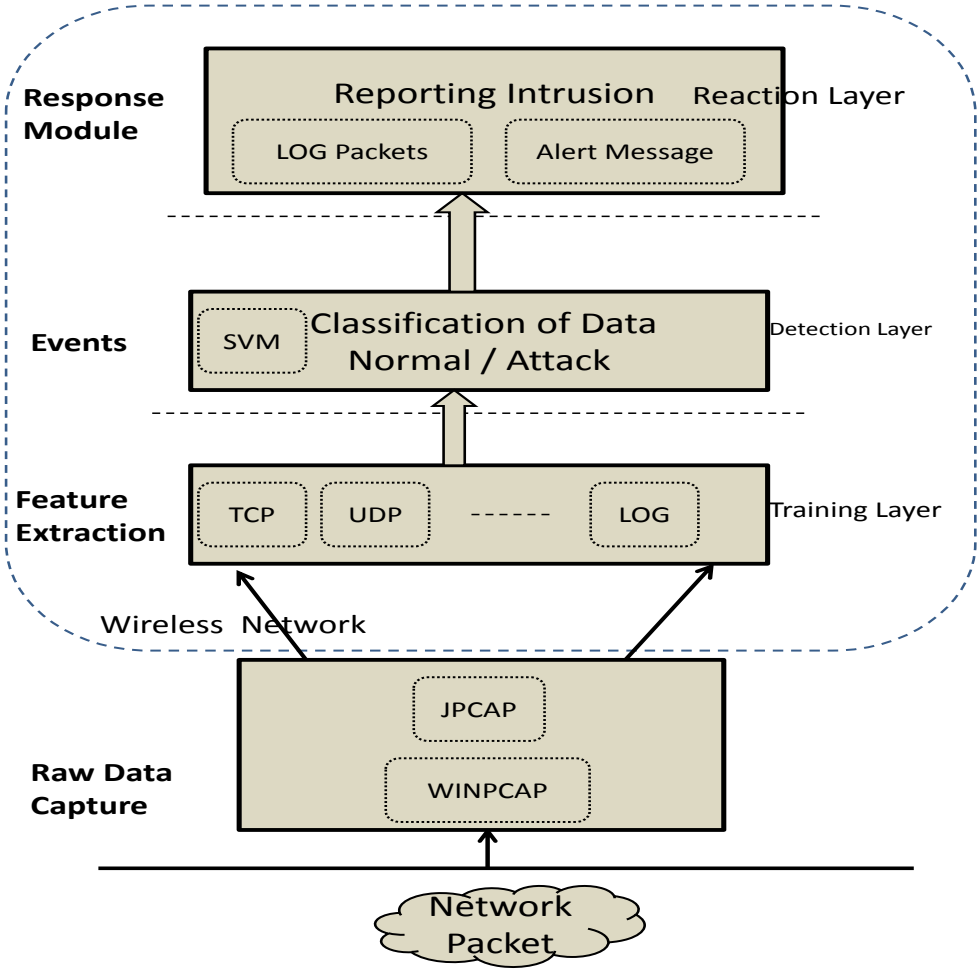

Fig. 1: Architecture of the Instruction Detection System

Scientia Africana, Vol. 13 (No.2), December 2014. Pp317-327
© College of Natural and Applied Sciences, University of Port Harcourt, Printed in Nigeria *ISSN 1118 - 1931*

Support Vector Machine Algorithm Analysis

An SVM maps linear algorithms into non-linear space. It uses a feature called, kernel function, for this mapping. Kernel functions like polynomial, radial basis function are used to divide the feature space by constructing a hyperplane. The kernel functions can be used at the time of training of the classifiers which selects support vectors along the surface of this function. SVM classify data by using these support vectors that outline the hyperplane in the feature space. This process will involve a quadratic programming problem, and this will get a global optimal solution. Suppose we have N training data points

{(x1, y1), (x2,y2), (x3, y3), ..., (xN , yN)}, where xi € Rd and yi €{+1,−1}.

Consider a hyper-plane defined by (w, b), where w is a weight vector and b is a bias.

The classification of a new object x is done with f(x) sign(w.x b) sign(α y (x x) b)

The training vectors xi occurs only in the form of a dot product. For each training point, there is a Lagrangian multiplier αi. The Lagrangian multiplier values αi reflect the importance of each data point. When the maximal margin hyper-plane is found, only points that lie closest to the hyper-plane will have αi > 0 and these points are called support vectors. All other points will have αi = 0. That means only those points that lie closest to the hyperplane, give the representation of the hypothesis/classifier. These data points serve as support vectors. Their values can be used to give an independent boundary with regard to the reliability of the hypothesis/classifier (Snehal et al, 2010). Fig 2 shows the graph of captured data.

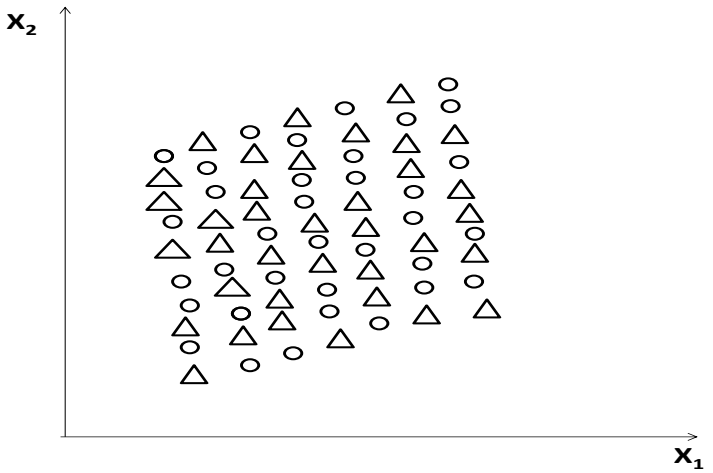

Fig. 2: A graph of all the packets captured before separation or classification

When the algorithm was applied, classification was done and the packets were separated into an attack and a normal packet. Fig 3 shows the classification of the packets.

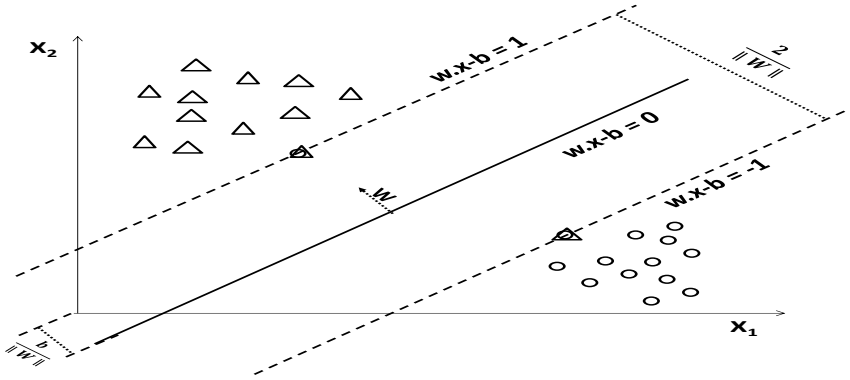

Fig. 3: A graph showing classified packets into two classes of normal and attack.

IMPLEMENTATION AND RESULTS

We used Java programming language in the development and implementation of an intrusion detection system based on support vector machine algorithm. The implementation of SVM intrusion detection system has two phases: training (learning) and testing (detection). The two phase; training and testing was performed on Wireless Local Area Network in a real time environment, for each TCP/IP connection, 10 various quantitative and qualitative features were extracted. During training the system, we click on the icon Anomaly IDS .jar in the figure 4. This activates the applications graphic user interface where we selected the network interface to use. All the four systems monitored were on the same wireless access point, the SVM model was initiated by clicking on SVM training button in the SVM training section on the GUI in figure 5. As the training progresses, we noticed the network packets were captured with the help of the packet sniffer, important features were selected and extracted, and rules are generated from the SVM training guiding the normal profile activities. These captured packets that were trained was finally subjected to testing to determine the effectiveness of the system developed.

Figure 4: An interface of the developed IDS showing different icons

Scientia Africana, Vol. 13 (No.2), December 2014. Pp317-327
 ISSN 1118 - 1931

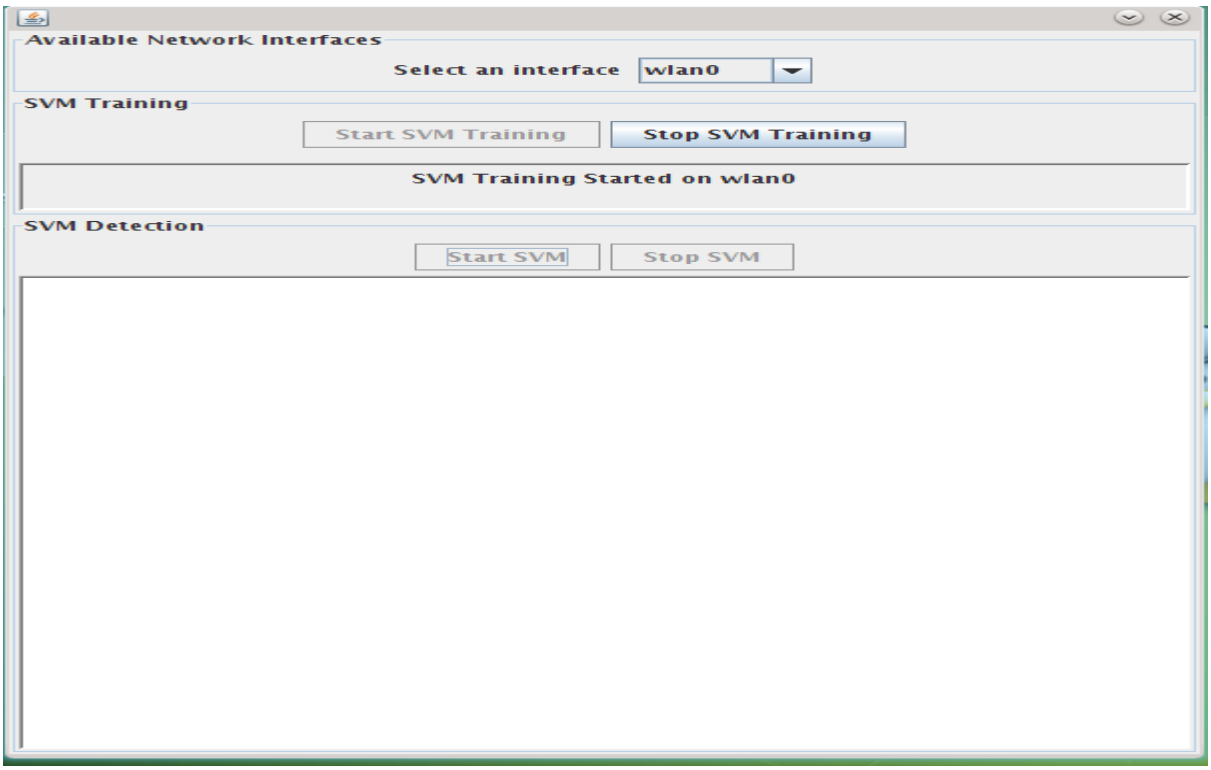

Figure 5: An interface showing the section of SVM modules

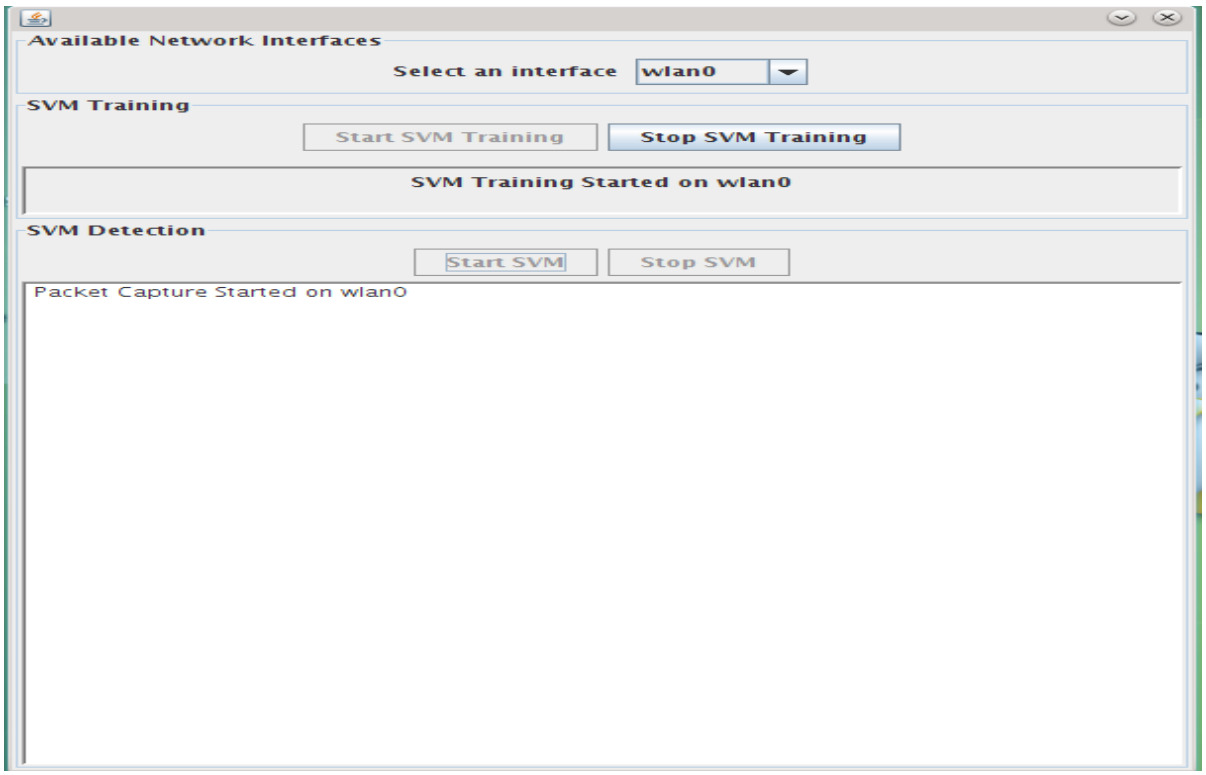

Figure 6: An interface showing packets capture and subjected to training

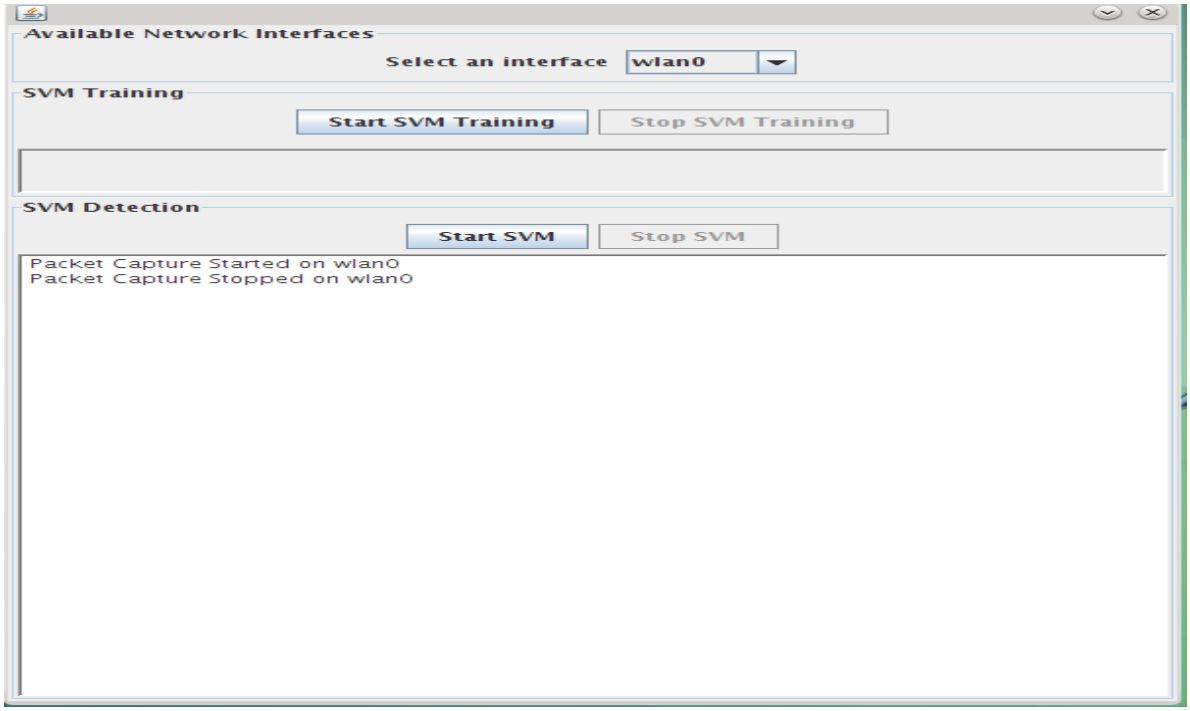

Figure 7: An interface indicating the completion of the training

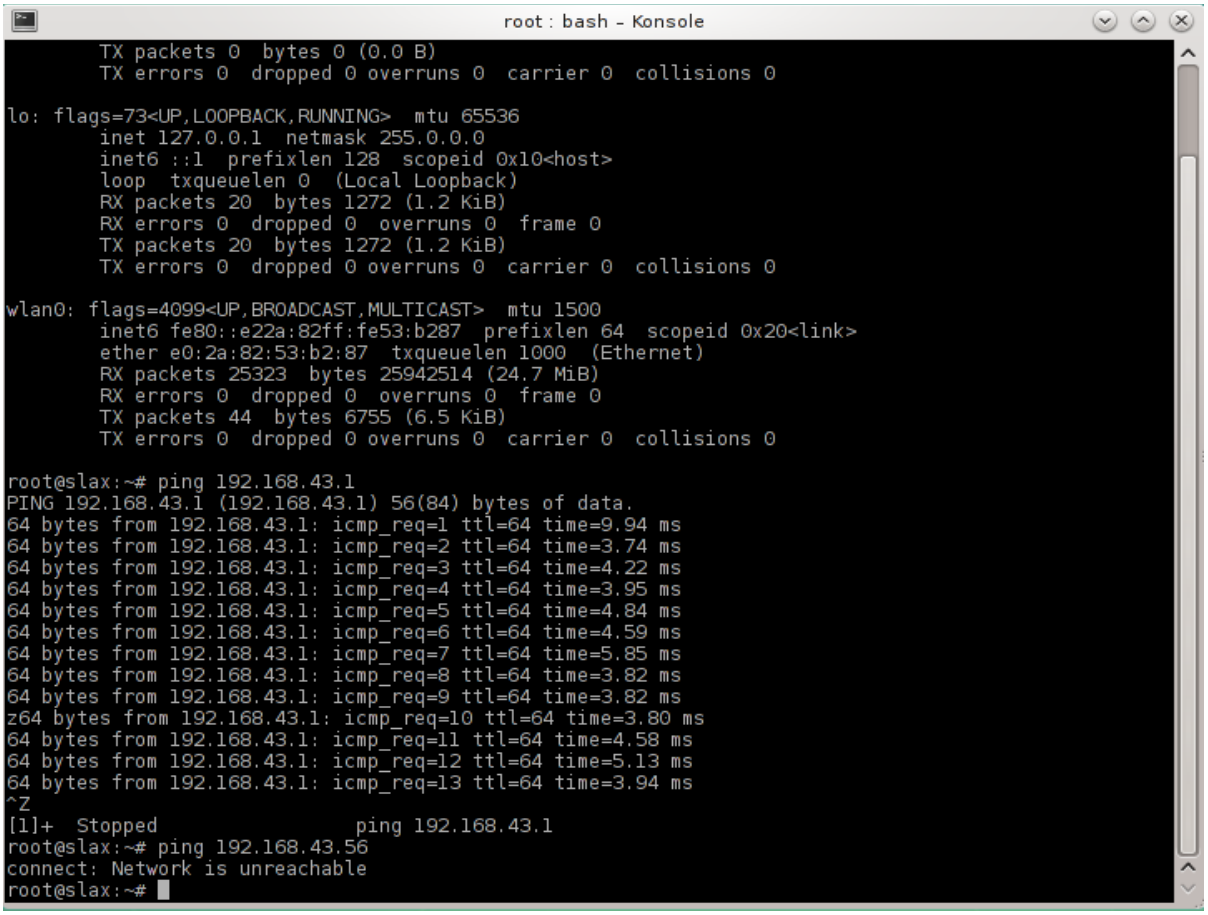

Figure 8: An interface showing an attack model

Scientia Africana, Vol. 13 (No.2), December 2014. Pp317-327
© College of Natural and Applied Sciences, University of Port Harcourt, Printed in Nigeria *ISSN 1118 - 1931*

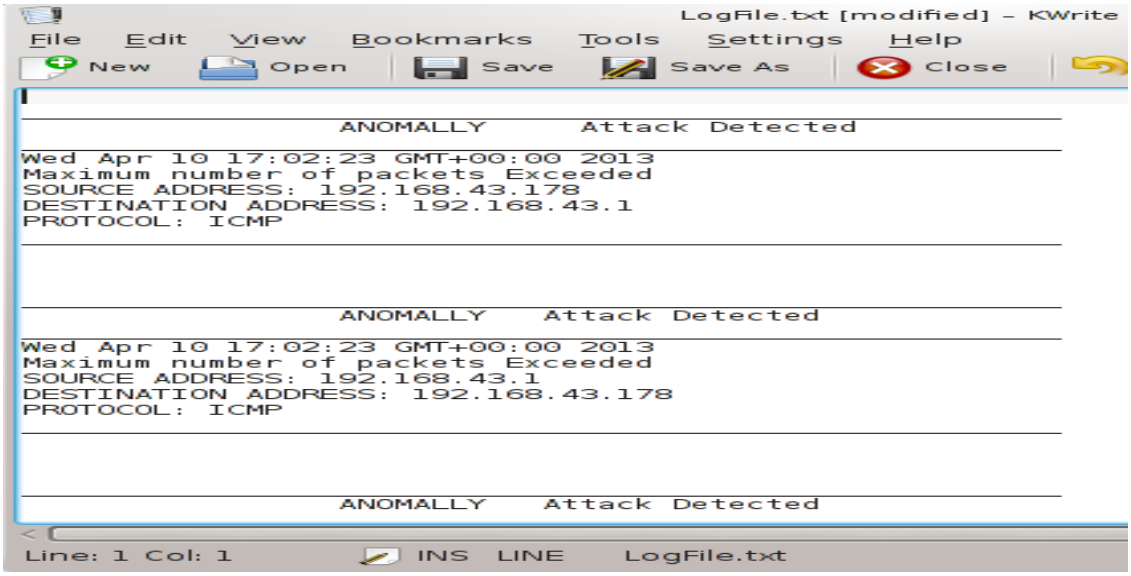

Figure 9: An interface showing result of detected attacks by IDS

Table 2: A summary of the experimental result

Class	Normal	Attack
Normal	44013.00 (TN)	98.00 (FP)
Attack	122 (FN)	10864.0 (TP)

If Detection Accuracy (DA) is 100 % if there is no misclassification of data. Calculating for different percentage of misclassification and subtracting it from the 100% expected gave us our system detection rate.

$\frac{FP}{FP+TN} \times 100 = \frac{98.00}{44,013+98} \times 100 =$ 0.22 Detection Accuracy = 100 – 0.22 = 99.78%

$\frac{FN}{FN+TP} \times 100 = \frac{122.00}{122.00+10864} \times 100 =$ 1.11 Detection Accuracy = 100-1.11 = 98.89%

From our results it shows our system classification is very high, that is looking at 99.78% and 100% for misclassification.

Table 3: showing percentage of normal and attacks from the application.

Class	Accuracy
Normal	99.78
Attack	98.89

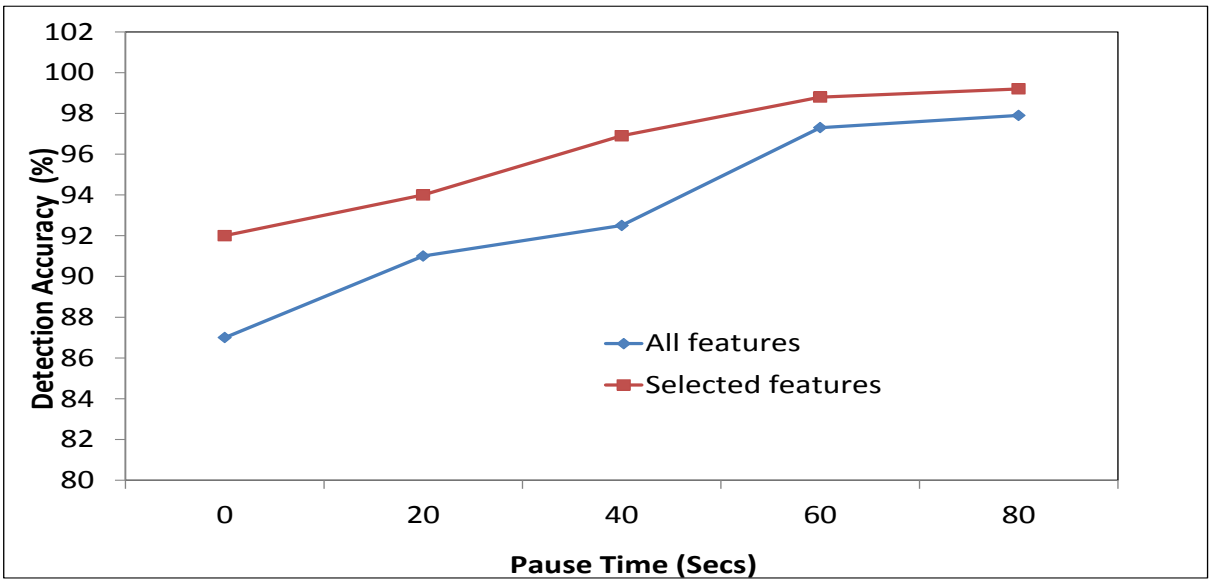

Fig 10: A graph showing all the features and selected features

DISCUSSION

The experimental results shown in fig 9 depicts the results obtained from testing the intrusion detection system developed, the system was tested with 5500 packets that has been trained, 44013 datasets were classified as normal packets and 10864 was classified as attacks while few were misclassified. This indicated low false alarm rate, detection accuracy was used as a metrics for evaluating the efficiency of the application developed. Table 2 and 3 shows the results obtained as 99.78% and 98.89% for the detection performance of application developed for normal and attacks when tested with 5500 packets. In the paper of Scherer, et al (2010), Using SVM and Clustering Algorithms in IDS Systems, the average success rate of classification was between 91.228% and 98.998% as against our success rate of 99.78% and 98.89%. Fig 10 shows a graph of all the features and selected features which buttress our augment on the efficiency of the developed system.

In this paper, we have been able to develop an improved intrusion detection system by using a classification algorithm, support vector machine approach and a hybrid force alarm detection methods. The test result shows that the intrusion detection system approach used improved the detection efficiency by reducing or eliminating false positives and false negatives of IDS. The hybrid system produced a better result of intrusion detection and has the advantage of a new attack signatures been detected and can be used to write signature in other to update the database of other intrusion detection system that is based on the misused approach alone.

REFERENCES

Anderson, J. R. (1989). A Theory of the Origins of Human Knowledge. Artificial Intelligence, Elsevier Science Publisher 40, 313–352.

Bakar A .A, Z. A. Othman, A. R. Hamdan, R. Yusof, and R. Ismail, (2008). "An Agent Based Rough Classifier for Data Mining," in Proceedings of the

Scientia Africana, Vol. 13 (No.2), December 2014. Pp317-327
© College of Natural and Applied Sciences, University of Port Harcourt, Printed in Nigeria *ISSN 1118 - 1931*

8th International Conference on Intelligent Systems Design and Applications Kaohsiung City, Taiwan, N. (ISDA'08), 145–151.

Dorothy E. Denning (1987). An Intrusion Detection Model, IEEE Transaction on Software Engineering(TSE), 13, 2, 222-232.

ElisseeA., and Guyon I. (2003). "An Introduction to Variable and Feature Selection. Journal of Machine learning Research 3 1157-1182

Friedman, Linial N, Nachman M, Pe'er I. (2000). "Using Bayesian Networks to Analyze Expression Data". Journal of Computational Biology 7(3– 4), 601–620

Hsu C, Lin J, (2002). A Comparison of Methods For Multiclass Support Vector Machines. IEEE Trans. Neural Networks. 13, 415-425.

Ilgun K, Kemmerer R, and Porras P (1995). State Transition Analysis: A RuleBased Intrusion Detection Approach. IEEE Trans Software Eng, 21, 3, 181-199.

Scarfone, Karen, Mell, and Peter (2007). "Guide to Intrusion Detection and Prevention System (IDPS)

Snehal, A Mulay, Devale, P. R and Garje, G. V. (2010). Intrusion Detection System Using Support Vector Machine and Decision Tree, International Journal of Computer Applications, l 3. 3, 40-43.

Tweedale, A. Quteishat, C. Peng Lim, and Jain L. C., (2009). "A Neural Network-Based Multi-Agent Classifier System," Neurocomputing, 72, 1639–1647.

Vapnik V. N.(1995).The Nature of Statistical Learning Theory, Springer-Verlag,New York. NY.

W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog,(2009). "Attribute Normalization in Network Intrusion Detection," in Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms, and Networks (I-SPAN'09), IEEE, Kaohsiung City, Taiwan, 448–453.

Vapnik. V. N. (1995) The nature of statistical learning theory. Springer-Verlag, New York. NY,

Winkeler, J.R. (1990). "A UNIX Prototype for Intrusion and Anomaly Detection inSecure Networks," The Thirteenth National Computer Security Conference, Washington, DC., pages 115–124.

Yongguang Z., Wenke L (2000). "Intrusion Detection in Wireless Ad- Hoc Networks",Proceedings of the Annual International Conference on Mobile Computing and Networking, MobiCom . 275-283.

Zaki, Mohammed J. (2001). SPADE: An Efficient Algorithm for Mining Frequent Sequences, Machine Learning Journal, 42, 31–60

Zanero S., (2005) "Improving Self Organizing Map Performance for Network Intrusion Detection", International Workshop on Clustering High-Dimensional data and its applications, SDM 05 SIAM conference On Data Mining. 30-37.