Intrusion Detection System Using Comparative Machine Learning Algorithms

Ifeanacho Francis Ejimofor[1], Aminat Abiola Ajibola[2], Uchenna Igboeli[3], Musah Abdulmumini Yakubu[4]

[1,4]*Research Scholar , University of Abuja, Abuja-Nigeria*
[2]*Senior Lecturer, University of Abuja, Abuja-Nigeria*
[3]*Lecturer, University of Abuja, Abuja-Nigeria*

Abstract:- **Data security in a network-based computer system has become a major challenge in theworld today. With the high increase of network traffic, hackers and malicious users aredevising new ways of network intrusion. In other to address this problem, an intrusiondetection system (IDS) is developed which will detect attacks in a computer network. In thisresearch, the KDDCup99 Test datasets is analyzed using certain machine learningalgorithms (Bayes Net, J48, Random Forest, and Random Tree) to determine the accuracy ofthese algorithms by classifying these attacks into their various classes. A constructiveresearch methodology is adopted throughout this research. The experimental results showthat the Random Forest and Random Tree algorithms are the most effective inperforming the classification technique on the Test dataset. The experimental tool used wasWEKA which was used to perform a correlation-based feature selection on the dataset with aBest First search method, and the parameters used for the computation are Precision, Recalland F-measure.The result shows that random tree and random forest performs better when the average is taken in terms of Precision,Recall and F-measure.**

Keywords:- **Intrusion Detection System; Neural Network, Support Vector Machine, Supervised Learning; Machine learning; Internet Packets**

I. INTRODUCTION

Network intrusion have predominantly increased following the rapid growth of network or internet technologies in different areas of social networking, e-learning, e-business etc. this has made the security of data from malicious Hackers more challenging. An Intrusion Detection System is an application used for monitoring the network and protecting it from the intruder. With the rapid progress in the internet-based technology, new application areas forcomputer network have emerged [1]. An Intrusion Detection System (IDS) can be classified into Network-bases IDS, Host-based IDS and Application-Based IDS.

Network-based intrusion detection systems gather information directly from a network and performs auditing on the attacks in the network as packets travels in the network. This type of IDS grants users the privilege to specify its signature. The Host-based IDS on the other hand, views the signature of intrusion in the local system. For analysis they use host system's logging and other information. Host based handler is referred as sensor [2].

Application-based IDS on its part, checks the effective behavior and event of the protocol. The system or agent is placed between a process and group of servers that monitors and analyzes the application protocol between devices [3].

II. REVIEW OF RELATED WORK

In the work of [4], they demonstrated the use of a hybridized machine learning models with the expectation of showing the capability to the job of intrusion detection in a computer network. The research used the training and testing versions of the NSL-KDD datasets in other to illustrate the effectiveness of the model against known and unknown entries in the model. This work made use of Neural Network (NN) and Support Vector Machine (SVM) algorithms for the supervised learning, K-Means algorithm for the unsupervised learning and PCA and GFR for feature selection on the datasets. In a similar study,[5] used seven machine learning algorithms to perform a supervised technique on the NSL-KDD dataset using WEKA as their desired data mining tool. The algorithms used to carry out this experimental work are: PART, Bayes Net, IBK, Logistic, J48, Random Committee and Input Mapped. [6]gave an overview and the major importance as it relates to intrusion detection system (IDS). The study gave a general insight on the major types of intrusion detection system, the attack types, diverse domains, attack tools and IDS lifecycle. The works of [7]gave a different approach to the development of IDS generally. The researchers adopted the use of Apriori algorithm alongside the association rules to solve the intrusion detection issues. This research applied an evasion technique in other to detect new attacks using information gotten from the set of known attacks in the datasets.

III. METHODOLOGY AND DESIGN

In this research, the design methodology to be used is a hybridized methodology combining the Top-Down design approach and the object-oriented design approach. The purpose for the top-down design approach is to follow the TCP/IP 5 layer architecture in other to analyze the requirements of the system. The top-down approach gives an opportunity to troubleshoot a system when a layer in the TCP/IP suite is having a problem.

However, the object-oriented design technique can be used in the phases of the software development life cycle (Analysis, Design and implementation). The object-oriented technique is used to specify the classes and objects of a system and the relationship between them. This object design can be converted in applications using object-oriented languages. One of the major advantages of object oriented design is code reusability and scalability.

3.1. System Requirement Gathering

The requirement elicitation is a process in requirement engineering, In this research, the requirement gathering went through three stages: interviews with a client, use cases, questionnaires, user observations etc. This research specifically focused onUser requirement, Functional requirement and Nonfunctional requirement.

User Requirement

The user can be a stakeholder, organization or a client who have special concern in the system. They may be directly or indirectly influenced by the system. The user requirement specifications in this research are:

i. This system will be used by different organizations, law enforcement agencies and individuals
ii. The system will be a distributed system
iii. The system will be durable
iv. The user interface of the system will be friendly
v. The system should effectively handle operational errors

Functional Requirement

The functional requirements analysis will make use of the top-level functions emanating from the activities in the system. The requirements for this research include:

i. The login page should be able to differentiate between a hack attempt and an authorized access into the system
ii. The system should generate an access time graph each time a user or hack attempt is made
iii. The system should capture the location of the hacker

Non-functional Requirement

The nonfunctional requirements in this system are outlined:

i. Interoperability of the system
ii. The cost of implementing the system
iii. Configuration of the system

3.2 Design Methodology

The design methodology used in this research is a hybridized design method which combines the Top Down design approach and the Object-Oriented design approach. The purpose for the top down design approach is to follow the TCP/IP 5-layer architecture in other to analyze the requirements of the new system. The top down approach gives an opportunity to troubleshoot a system when a layer in the TCP/IP suite is having a problem. However, the object-oriented technique is used to specify the classes and objects of a system and the relationship between them.

3.3 System Architecture

System architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system [8]

3.4 Use Case Diagram

The use-case design of the proposed system is represented in figure 1.0

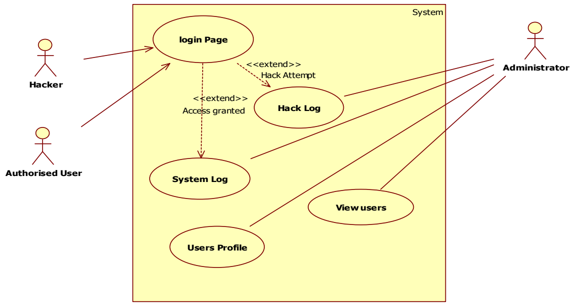


Figure 1.0

From the diagram in figure 1.0 show various actors as seen in table 1.0. The actor could be an authorized user, a hacker or system administrator who will only gain access through the login page (user case),the administrator would be able to view the user active on the system, the hack logs, user logs and the user profiles

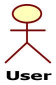
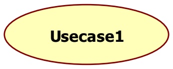
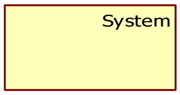
Symbol	Description
	**Actor:** An actor is an entity with behavior; this may be a person, computer system or an organization. Eg a customer.
	**Use case:** this is a single unit of work in a system. It is the collection of related success and failures scenarios that describes actors using a system.
	**System Boundary:** can be seen as an enclosure that illustrates the border between the entities that the use cases are representing (within the boundary) and the actors (outside the boundary).

Table 1.0

3.5 Class Diagram

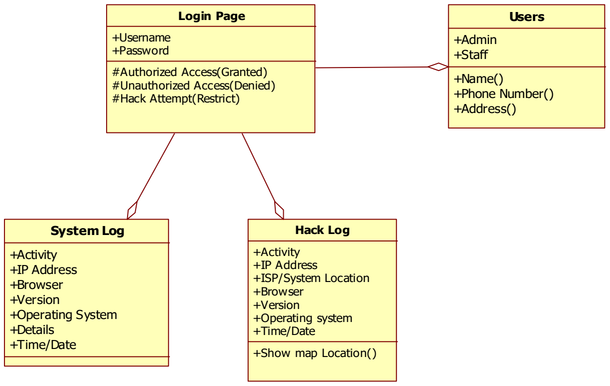


Figure 2.0

From the diagram in figure 2.0 there are four classes (Login page, users, system log and Hack log). The login page have username and password as the attributes of the class, the operations that is found in this class are granted, denied or restrict, and they are set to be protected. The Users class have two attributes (Admin and Staff class), each of them have their attributes (Name, phone number, address etc.) that helps to identify the user of the system.

The System log and the hack log classes both have almost the same attributes, the Hack log have the ISP/System location which captures the location of the IP address of where the hack is being lunched. The hack logs also have a show map location that points to the actual position the system or ISP Address from which the hack attempt was carried out. The classes (Users, system log and hack log) are connected to the login page class using an aggregation which is a more specific type of association.

3.6 Data Flow Diagram

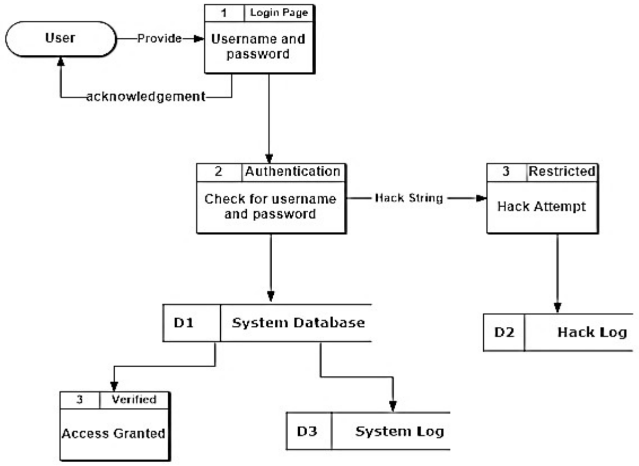


Figure 3.0

The diagram in figure 3.0 illustrates the flow of data in the system. The user supplies a username and password to gain access into the system after which an acknowledge message is sent to the user specifying if the login was successful or denied. The user credentials goes through an authentication process to determine if it's a hack attempt or a valid user. This information is stored in hack attempt log or system log respectively.

IV. DATA PRESENTATION, ANALYSIS AND DISCUSSION

Hack attempt log table

The diagram in table 2.0 captures the intruder's IP address, location on the map, and other details that can be captured over the internet. It also shows the details of hack attempts made on the system.

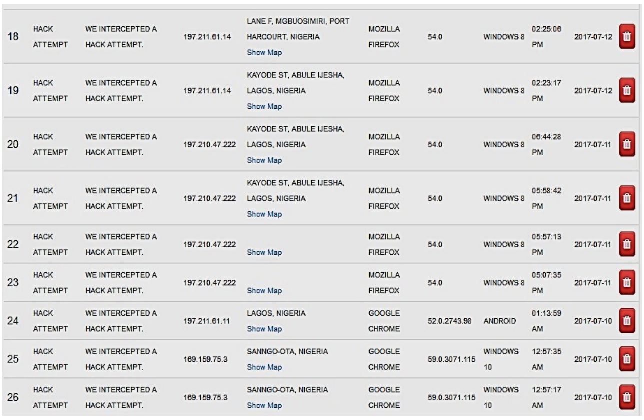
An intrusion detection system (IDS) is evaluated by the measure of accuracy, detection rate and F-measure. An intrusion detection system should have a very low false alarm.

Precision is the percentage of the total number of attacks that are properly detected. Detection Rate or Recall is described as the number of attacks detected by the proposed technique to the total number of attacks truly there [9]while True Positive (TP) is the number of connections that were correctly classified as an intrusion

$$\text{Accuracy(Precision)} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

While False Positive (FP) is the number of intrusion connections that were incorrectly classified as normal, False Negative (FN) shows the number of normal connections that were incorrectly classified as intrusion

$$\text{F} - \text{Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$



18	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	197.211.61.14	LANE F, MGBUOSIMIRI, PORT HARCOURT, NIGERIA Show Map	MOZILLA FIREFOX	54.0	WINDOWS 8	02:25:06 PM	2017-07-12	🗑
19	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	197.211.61.14	KAYODE ST, ABULE IJESHA, LAGOS, NIGERIA Show Map	MOZILLA FIREFOX	54.0	WINDOWS 8	02:23:17 PM	2017-07-12	🗑
20	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	197.210.47.222	KAYODE ST, ABULE IJESHA, LAGOS, NIGERIA Show Map	MOZILLA FIREFOX	54.0	WINDOWS 8	06:44:28 PM	2017-07-11	🗑
21	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	197.210.47.222	KAYODE ST, ABULE IJESHA, LAGOS, NIGERIA Show Map	MOZILLA FIREFOX	54.0	WINDOWS 8	05:58:42 PM	2017-07-11	🗑
22	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	197.210.47.222	Show Map	MOZILLA FIREFOX	54.0	WINDOWS 8	05:57:13 PM	2017-07-11	🗑
23	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	197.210.47.222	Show Map	MOZILLA FIREFOX	54.0	WINDOWS 8	05:07:35 PM	2017-07-11	🗑
24	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	197.211.61.11	LAGOS, NIGERIA Show Map	GOOGLE CHROME	52.0.2743.98	ANDROID	01:13:59 AM	2017-07-10	🗑
25	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	169.159.75.3	SANNGO-OTA, NIGERIA Show Map	GOOGLE CHROME	59.0.3071.115	WINDOWS 10	12:57:35 AM	2017-07-10	🗑
26	HACK ATTEMPT	WE INTERCEPTED A HACK ATTEMPT.	169.159.75.3	SANNGO-OTA, NIGERIA Show Map	GOOGLE CHROME	59.0.3071.115	WINDOWS 10	12:57:17 AM	2017-07-10	🗑

Table 2.0

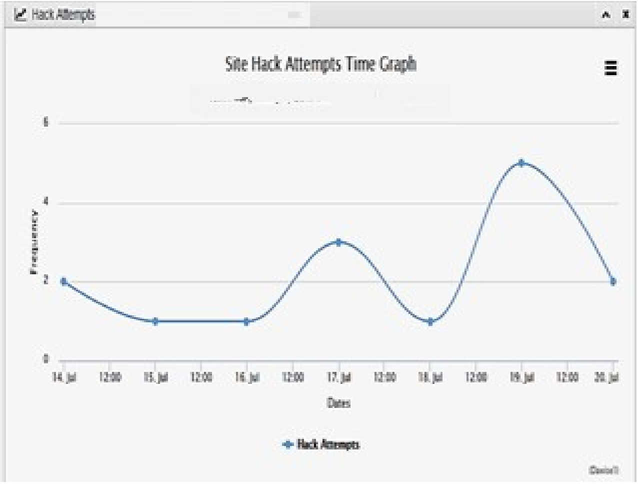


Figure 4.0

The graph in figure 4.0 shows the period with the highest frequency of attacks. From the graph it is seen that in the month of July they were more unauthorized users trying to get access to the system

Presentation **of Results**

In other to carry out the experimental analysis in this research, the tool used is WEKA 3.8 (Waikato Environment for Knowledge Analysis). This is open source machine learning scripting software developed in Java by the Waikato University, New Zealand [10].

	BayesNet (%)	J48 (%)	RandomForest (%)	RandomTree (%)
Precision	86.1	96.2	97.7	97.4
Recall	90.6	86.1	87.5	88.8
F-Measure	87.2	90.3	91.8	92.3

Table 3.0

From the result in Table 3.0 it is seen that BayesNet algorithm has the least precision,J48algorithm has least recall rate ,while BayesNet algorithms. When the average is take it is seen that Random Forest and Random Tree algorithms outperforms the BayesNet and J48 Algorithims

V. CONCLUSIONS AND RECOMMENDATIONS

Conclusion

Data and information security is one of the major challenges faced today due to the tremendous growth of internet packets and the incessant increase of malicious attacks, it is only predominant that intrusion detection system (IDS) will go a long way to curb and checkmate the rate at which these attackers gain access into network systems. This research developed an intrusion detection system (IDS) that will control access and detect attacks. The intrusion detection system (IDS) was developed in other to detect attacks that are launched by hackers in a network and the location of the hacker on a map.

With the use of machine learning algorithms, classification analysis is performed on the Testing dataset of the KDDCup99 datasets which was used to train the system. This research also described the various attack types and there classes, using feature engineering to remove the irrelevant features in the datasets in other to improve the classification accuracy. On a weighted average, it is evident that two (2) algorithms (RandomForest and RandomTree) performed better than the others as they are above 97%.

Recommendations

Based on the knowledge gained from this research and the tremendous value it will add to academic research and to the security agencies, there are recommendations that will be considered for future work:

1. Future research should consider other machine learning algorithms to ascertain other efficient ways to perform the classification technique on the datasets.
2. It is recommended that further research should be carry out on other parameters that can further improve the accuracy of detection

REFERENCES

[1] Kabiri, P. and A.A. Ghorbani, *Research on intrusion detection and response: A survey.* IJ Network Security, 2005. **1**(2): p. 84-102.

[2] Bace, R., *An introduction to intrusion detection and assessment for system and network security management.* ICSA Intrusion Detection Systems Consortium Technical Report, 1999.

[3] Ugochukwu, C.J. and E. Bennett, *An intrusion detection system using machine learning algorithm.* International Journal of Computer Science and Mathematical Theory, 2018. **4**(1): p. 39-47.

[4] Abubakar, A. and B. Pranggono. *Machine learning based intrusion detection system for software defined networks*. in *2017 Seventh International Conference on Emerging Security Technologies (EST)*. 2017. IEEE.

[5] Noureldien, N. and I. Yousif, *Accuracy of machine learning algorithms in detecting DoS attacks types.* Science and Technology, 2016. **6**(4): p. 89-92.

[6] Vijayarani, S. and M. Sylviaa, *Intrusion detection system-a study.* IJSPTM, 2015. **4**(1): p. 31-44.

[7] Nalavade, K. and B. Meshram, *Finding Frequent Itemsets using Apriori Algorithm to Detect Intrusions in Large Dataset.* International Journal of Computer Applications & Information Technology, 2014. **6**(1): p. 84-92.

[8] Jaakkola, H. and B. Thalheim. *Architecture-Driven Modelling Methodologies*. in *EJC*. 2010.

[9] Modi, U. and A. Jain, *An improved method to detect intrusion using machine learning algorithms.* Informatics Engineering, an International Journal (IEIJ), 2016. **4**(2): p. 17-29.

[10] Hall, M., et al., *The WEKA data mining software: an update.* ACM SIGKDD explorations newsletter, 2009. **11**(1): p. 10-18.