INTRUSION DETECTION SYSTEM USING MACHINE LEARNING ALGORITHMS IN CLOUD ENVIRONMENT

PROJECT GUIDE : Dr. S BOSE

Team Members :

1. Araventh M - 2019103508
2. Gangaraju Tharun - 2019103520
3. Sourabh Sonny - 2019103064
4. Raj Kumar J - 2019103564

INTRODUCTION :

- An intrusion detection system (IDS) is a monitoring tool that keeps track of suspicious activity and sends out warnings when it finds something suspicious.
- To address this issue, machine learning algorithms are increasingly being used in IDSs to improve their accuracy and reduce false positives.
- The proposed hybrid IDS uses a combination of supervised and unsupervised learning algorithms such as SVM, LSTM for intrusion detection.
- The IDS is evaluated using the CICIDS2017 dataset and the results indicate that the proposed IDS is able to detect intrusions with a high degree of accuracy

OVERALL OBJECTIVE :

- The proposed system should be able to detect, prevent, and respond to cyberattacks by utilizing ML techniques.
- The system will be able to identify malicious network traffic and alert the user in case of any suspicious activities.
- The system will be able to analyze the network traffic, identify any malicious behavior, and take appropriate action such as blocking the malicious traffic or alerting the user.
- The system should also be able to detect and respond to zero-day attacks.

LITERATURE SURVEY :

S.No.	Author, Publication, Year, Title	Proposed Work	Advantages	Disadvantages
1.	W. Wang, X. Du, D. Shan, R. Qin and N. Wang, "Cloud Intrusion Detection Method Based on Stacked Contractive Auto-Encoder and Support Vector Machine," in *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 1634-1646, 1 July-Sept. 2022.	Support Vector Machine	SCAE(feature extraction)	Cannot efficiently detect unknown attacks
2.	F. I. Shiri, B. Shanmugam and N. B. Idris, "A parallel technique for improving the performance of signature-based network intrusion detection system," 2011 IEEE 3rd International Conference on Communication Software and Networks, 2011, pp. 692-696.	Parallel Processing	Reduce Process Time	Cannot detect unknown attacks

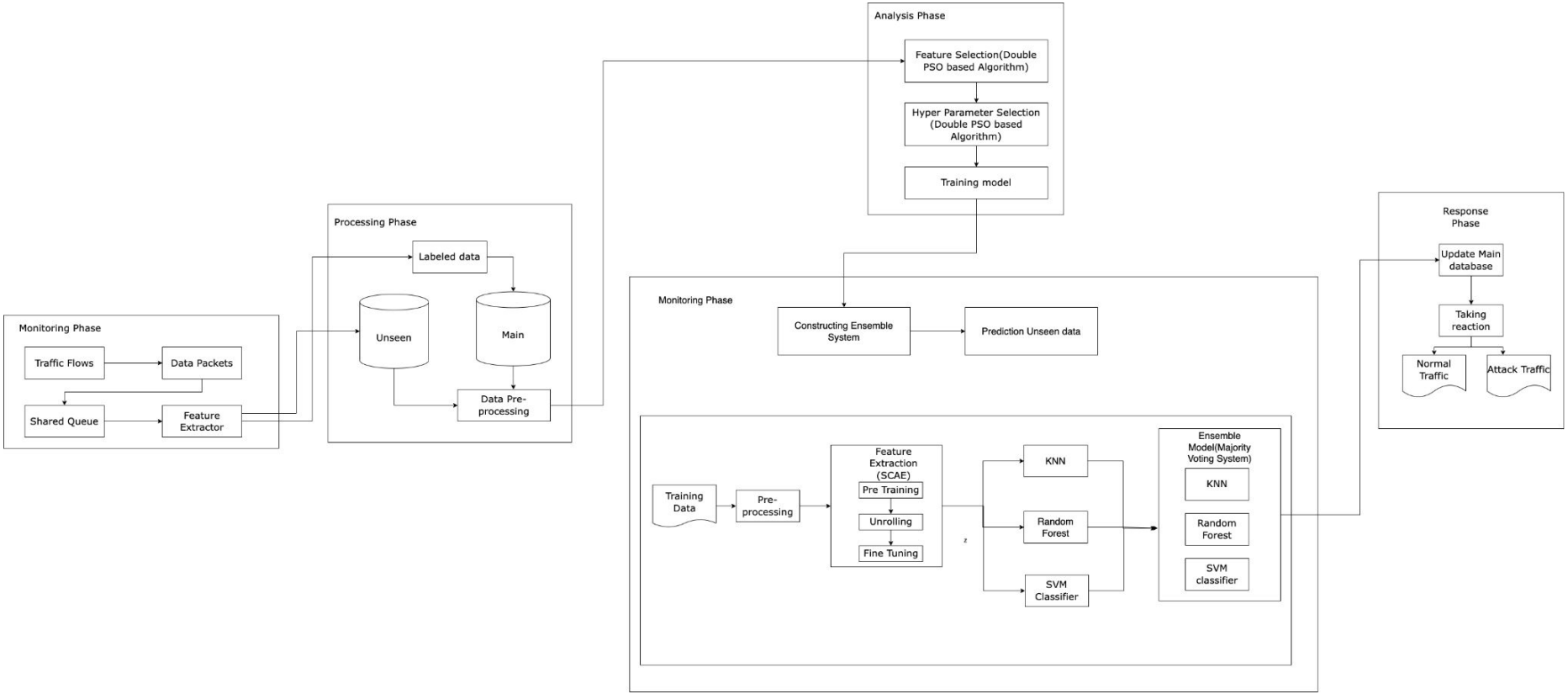
S.No.	Author, Publication, Year, Title	Proposed Work	Advantages	Disadvantages
3.	A. M. Vartouni, S. S. Kashi and M. Teshnehlab, "An anomaly detection method to detect web attacks using Stacked Auto-Encoder," 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), 2018, pp. 131-134.	Isolation Forest	Detection of unknown attacks	High false alarm rate
4.	A. Kannan, G. Q. Maguire Jr., A. Sharma and P. Schoo, "Genetic Algorithm Based Feature Selection Algorithm for Effective Intrusion Detection in Cloud Networks," 2012 IEEE 12th International Conference on Data Mining Workshops, 2012, pp. 416-423.	Genetic Feature Selection	Low false alarm rate	Features are removed without any extractions.

S.No.	Author, Publication, Year, Title	Proposed Work	Advantages	Disadvantages
5.	H. A. Kholidy and F. Baiardi, "CIDS: A Framework for Intrusion Detection in Cloud Systems," 2012 Ninth International Conference on Information Technology - New Generations, Las Vegas, NV, USA, 2012, pp. 379-385	P2P Network Architecture	Flexibility and Scalability	Not sufficiet for detecting large scale attacks
6.	A. Javadpour, S. Kazemi Abharian and G. Wang, "Feature Selection and Intrusion Detection in Cloud Environment Based on Machine Learning Algorithms," 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), Guangzhou, China, 2017, pp. 1417-1421F	Neural Network, Fuzzy Logic	Suitable for Qualitative features	Low Flexibility

S.No.	Author, Publication, Year, Title	Proposed Work	Advantages	Disadvantages
7.	G. Kene and D. P. Theng, "A review on intrusion detection techniques for cloud computing and security challenges," 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 2015, pp. 227-232	Hybrid based detection	Accurate detection of known attacks	Can’t detect unknown attacks
8.	C. -C. Lo, C. -C. Huang and J. Ku, "A Cooperative Intrusion Detection System Framework for Cloud Computing Networks," 2010 39th International Conference on Parallel Processing Workshops, San Diego, CA, USA, 2010, pp. 280-284	Majority Vote Method	Better accuracy	Less number of attacks detected

S.No.	Author, Publication, Year, Title	Proposed Work	Advantages	Disadvantages
9.	M. Ficco, L. Tasquier and R. Aversa, "Intrusion Detection in Cloud Computing," 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, Compiegne, France, 2013, pp. 276-283	Artificial Neural Network	Central correlation system to send alerts	Less Accuracy
10.	U. Oktay and O. K. Sahingoz, "Proxy Network Intrusion Detection System for cloud computing," 2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), Konya, Turkey, 2013, pp. 98-104	Deep learning algorithms	Better at hardware usage	Time taking

BLOCK DIAGRAM :



DETAILS OF MODULES :

MONITORING PHASE :

- It carries out two distinct tasks: feature extraction and traffic flow capture.
- The collected traffic flows is stored immediately in the shared queue.
- Further, feature extractor handles every data packets in the shared queue
- Then it stores all of the extracted data packet characteristics in a data record or tuple and passes it to the next module.

PROCESSING PHASE :

- It tries to assign the already seen data into the proper class label
- It will be determined by using the Signature based detection(Decision Tree)
- Stores the labeled data into the main database
- There will be unseen database where the unseen data will store, which contains unlabeled datasets

ANALYSIS PHASE :

- Labeled and seen data will come as an input to this phase
- Performs pre-training and training phases of the deep learning models
- Executes the double PSO based algorithm for both feature and hyperparameter selection
- After applying the upper level of the double PSO based algorithm we will get optimal feature subset as the output
- After applying the lower level we will get optimal hyper parameter vector
- After training the model we will get the predicted output for the labeled and seen data

MONITORING PHASE :

- The prediction module, which performs two sequential tasks
- The first step is to build the bagging ensemble system utilising the previously trained KNN, Random Forest, and SVM models.
- The processing module then obtains a duplicate of the unseen database and does data preprocessing on it as well.
- The prediction module's second purpose is to successively choose a data record from the preprocessed unseen database.
- Following that, the selected data record is tested using the bagging ensemble system, and the majority voting engine's final judgement is sent to the response module.

RESPONSE PHASE :

- After combining the base model and the input dataset the result will be either Normal or Attack traffic
- The result will be updated into the main database.

DATASET :

- Canadian Institute for Cybersecurity Intrusion Detection Systems 2017 (CICIDS2017) is a fully labelled dataset with 80 network traffic characteristics derived from raw pcap data using the CICFlowMeter. Furthermore, the traits are retrieved and estimated for both benign and malicious flows depending on various applications and network protocols.
- They carried out 20 distinct attack types in order to cover a wide range of frequent attack scenarios, which may be classified into seven broad attack families or categories, namely, Brute Force, Heartbleed, Botnet, Denial of Service (DoS), DDoS, Web assault, and Infiltration.

- Here, in this project we are going to use 40 optimal features for this IDS. Some of those are Source Ip, Destination Ip, Source port, Destination port, Transport protocol, No of Submitted bytes, transmitted packets, Flow duration, Active mean etc..

Category		Total	Total(-rows with lack info)	Training	Test
BENIGN	BENIGN	2273097	2271320	20000	20000
DOS	DDoS	128027	128025	2700	3300
	DoS slowloris	5796	5796	1350	1650
	DoS Slowhttptest	5499	5499	2171	1169
	DoS Hulk	231073	230124	4500	5500
	DoS GoldenEye	10293	10293	1300	700
	Heartbleed	11	11	5	5
PortScan	PortScan	158930	158804	3808	4192
Bot	Bot	1966	1956	936	624
Brute-Force	FTP-Patator	7938	7935	900	1100
	SSH-Patator	5897	5897	900	1100
Web Attack	Web Attack-Brute Force	1507	1507	910	490
	Web Attack-XSS	652	652	480	160
	Web Attack-SQL Injection	21	21	16	4
Infiltration	Infiltration	36	36	24	6
Total Attack		471454	470365	20000	20000
Total		2830743	2827876	40000	40000

PERFORMANCE MEASURES :

THRESHOLD METRICS :

- Classification Rate (CR)
- F Measure (FM)
- Cost Per Example (CPE)

RANKING METRICS :

- False Positive Rate (FPR)
- Detection Rate (DR)
- Intrusion Detection Capability (CID)
- Precision (PR)
- Area Under ROC Curve (AUC)

PROBABILITY METRICS :

- Root Mean Square Error (RMSE)

REFERENCES :

W. Wang, X. Du, D. Shan, R. Qin and N. Wang, "Cloud Intrusion Detection Method Based on Stacked Contractive Auto-Encoder and Support Vector Machine," in *IEEE Transactions on Cloud Computing*, vol. 10, no. 3, pp. 1634-1646, 1 July-Sept. 2022.

F. I. Shiri, B. Shanmugam and N. B. Idris, "A parallel technique for improving the performance of signature-based network intrusion detection system," 2011 IEEE 3rd International Conference on Communication Software and Networks, 2011, pp. 692-696.

A. M. Vartouni, S. S. Kashi and M. Teshnehlab, "An anomaly detection method to detect web attacks using Stacked Auto-Encoder," 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), 2018, pp. 131-134.

A. Kannan, G. Q. Maguire Jr., A. Sharma and P. Schoo, "Genetic Algorithm Based Feature Selection Algorithm for Effective Intrusion Detection in Cloud Networks," 2012 IEEE 12th International Conference on Data Mining Workshops, 2012, pp. 416-423.

H. A. Kholidy and F. Baiardi, "CIDS: A Framework for Intrusion Detection in Cloud Systems," 2012 Ninth International Conference on Information Technology - New Generations, Las Vegas, NV, USA, 2012, pp. 379-385

A. Javadpour, S. Kazemi Abharian and G. Wang, "Feature Selection and Intrusion Detection in Cloud Environment Based on Machine Learning Algorithms," 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), Guangzhou, China, 2017, pp. 1417-1421F

G. Kene and D. P. Theng, "A review on intrusion detection techniques for cloud computing and security challenges," 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 2015, pp. 227-232

C. -C. Lo, C. -C. Huang and J. Ku, "A Cooperative Intrusion Detection System Framework for Cloud Computing Networks," 2010 39th International Conference on Parallel Processing Workshops, San Diego, CA, USA, 2010, pp. 280-284

M. Ficco, L. Tasquier and R. Aversa, "Intrusion Detection in Cloud Computing," 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, Compiegne, France, 2013, pp. 276-283

U. Oktay and O. K. Sahingoz, "Proxy Network Intrusion Detection System for cloud computing," 2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), Konya, Turkey, 2013, pp. 98-104