IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EPOERA — An Ensemble Based Measurement Learning Framework for Proactive Cyber Attack Detection in Cyber Range Environments

G. Divya[1]	Dr P. Prashanth Kumar[2]	Dr V. Anantha Krishna[3]	Dr U. Srilakshmi[4]
M.Tech[1]	Associate Professor[2]	Professor[3]	Professor & HOD[4]
Department of Computer Science and Engineering[1234]			
Sridevi Women's Engineering College, Hyderabad, Telangana, India[1234]			

Abstract **-** There was no centralized planning for the rapid acceleration of digital transformation across banking and educational infrastructures. The increasing severity of cybercrime revenue presents new engineering challenges. We suggest a framework based on the use of measurement learning in controlled cyber range settings to create baseline normality patterns allowing for proactive detection. We combine network- and host-based measurements in a SOC context. For our modelling, we used an ensemble multi-classifier (Decision Tree, Logistic Regression, Support Vector Machine (SVM), and Gradient Boosting).

This interaction will tackle the sluggish response of existing Artificial Neural Network (ANN) solutions in cloud data analysis by concentrating high innovate edge computing in embodied agent level upon willingness to utilize joker function. When examined experimentally, the system performs well at recognizing IOCs such as unauthorized authentication attempts and malicious pivoting commands. Let us see the implementation which uses ELK-Stack for centralized logs collection, organisation. Because the underlying dataset is granular with source strings and geographic coordinates, our framework excels at a high accuracy.

Keywords: Cyber Range; Machine Learning; Network Intrusion Detection; SOC; Measurement Learning, ELK-Stack.

I. INTRODUCTION

Society the world over is coming together to technology-led infrastructures. Digital transformation increases boundaries for organizations, companies, and educational institutions, yet this transition was not predicted or prepared. Cybercrime plots simultaneously spring into existence. The reality from recent reports is that cybercrime now operates on a competitive level with real-world crime regarding incident volume and earnings. It is time to find a threat now. Cyber defense is mandatory to secure all cyber real estate because old tech will falter. Classic IDS systems generally test the performance of only two supported machine learning algorithms, namely the SVM and the ANN architecture. These systems are hampered by complex challenges where high data volumes associated with the cloud make their classification abilities perform lower. Current approaches never use network monitoring or implement ELK Stack prerequisites or endpoint detection response (EDR). They never progress the reach of surveillance through constant extraction from the real-time layer; hence, the so-called ground-truth is never gained.

Our technique is based on measurement learning within controlled cyber range environments. The system keeps a close watch on all network traffic, system log data, and user interactions. We identify baselines representing normality and teach machines how to behave normally. Then we find anything which seems different and poses a threat. This process is repetitive. An effective approach in building a dynamic security environment ensures the framework addresses particular environments such as those belonging to industrial entities or banking institutions. By analysing normal patterns, anomalies representing threats can be detected

with the help of various attributes derived directly from the Cyber Range telemetry. These 14 attributes—**rid, datetime, host, src, proto, type, spt, dpt, srcstr, cc, country, locale, lat, and long**—accurately predict potential malicious behaviour.

This project proposal comes within the scope of the Security Operations Centre (SOC), where we address simultaneously host and network measurements that researchers used to detect tactics, techniques, and procedures. Using such measures allows identifying activities of compromise that are currently happening. The approach adopted leverages the **ELK Stack (Elasticsearch, Logstash, Kibana)** for consolidating in a single platform the data acquired using specific logging strategies. This ensures the accuracy of measurements and high performances in log ingestion. We show high efficiency in this regard and correlate it with attackers' intention detection. We established a multi-classifier scheme with the help of **Decision Tree, Logistic Regression, SVM, and Gradient Boosting**. The issue in existing models was that the monitoring dependency has been absent. For prediction, the output helps detect whether a cyber-attack has been observed based on accuracies obtained. Measuring consistently helps to improve the accuracies of the models with the passing of time.

We demonstrate specific traces of attacker activities, like guessing passwords or moving laterally through routing commands to pivot. We make three core technical contributions in this research work. First, we build a proactive model addressing the issue of low latency of cloud data due to legacy ANN processes. Second, we use the ELK-Stack to show attacker's traces, such as unauthorized command execution and application usage. Third, we build a host-level telemetry dataset to create ground truth in a complex environment and make public IP addresses and binaries footprints available. With the enhancement, capacity will be built in the future by providing cloud datasets.

1.1. OBJECTIVE

1.1.1. GENERAL OBJECTIVE

The objective of this research work is to design and implement a framework which helps in detecting cyber-attacks on specific environments, be it an industrial enterprise, an educational institution or a banking enterprise.

1.1.2. SPECIFIC OBJECTIVE

In order to detect such cyber-attacks we measure, observe and analyze parameters of: •Network traffic (e.g. Packet size, type of traffic or protocol, connection time) •System logs (e.g. Logon attempts, error logs, amount of used resources) •User activity (e.g. Frequency of activity, any atypical activity). We will use ML to differentiate "normal" activity from suspicious, and train and test various models, considering their respective parameters (accuracy, recall, precision, number of false positives). ML algorithms with better parameters will be used in the system design. This system provides security managers and organizations with a tool which helps in making decision for mitigation and further action to improve security. It may help organizations reduce downtime, increase awareness of threats and protect their intellectual property and data.

1.2. Scope

The scope of this project can be divided into following parts: Immediate scope: In the immediate scope of the project, it would be detecting cyberattacks in the boundary of the cyber range or one specific organization or institution. Here we would be considering different datasets from a cyber range and using ML to classify them as normal or malicious traffic. Extended scope: In the extended scope, the system can be made to handle large scale systems using cloud and distribute the load for detecting across the boundaries of multiple organization using distributed systems and data storage.

Research scope: It opens scope of doing more research on how cyber ranges can be used with ML based detection.

More work can be done on incorporating different forms of ML like deep learning, federated learning, etc and on real time intrusion detection systems. Practical scope: The application of the project could be in organizations like banking sector, health sector, and education sectors which are vulnerable to attacks. This will improve the defense system of these organizations.

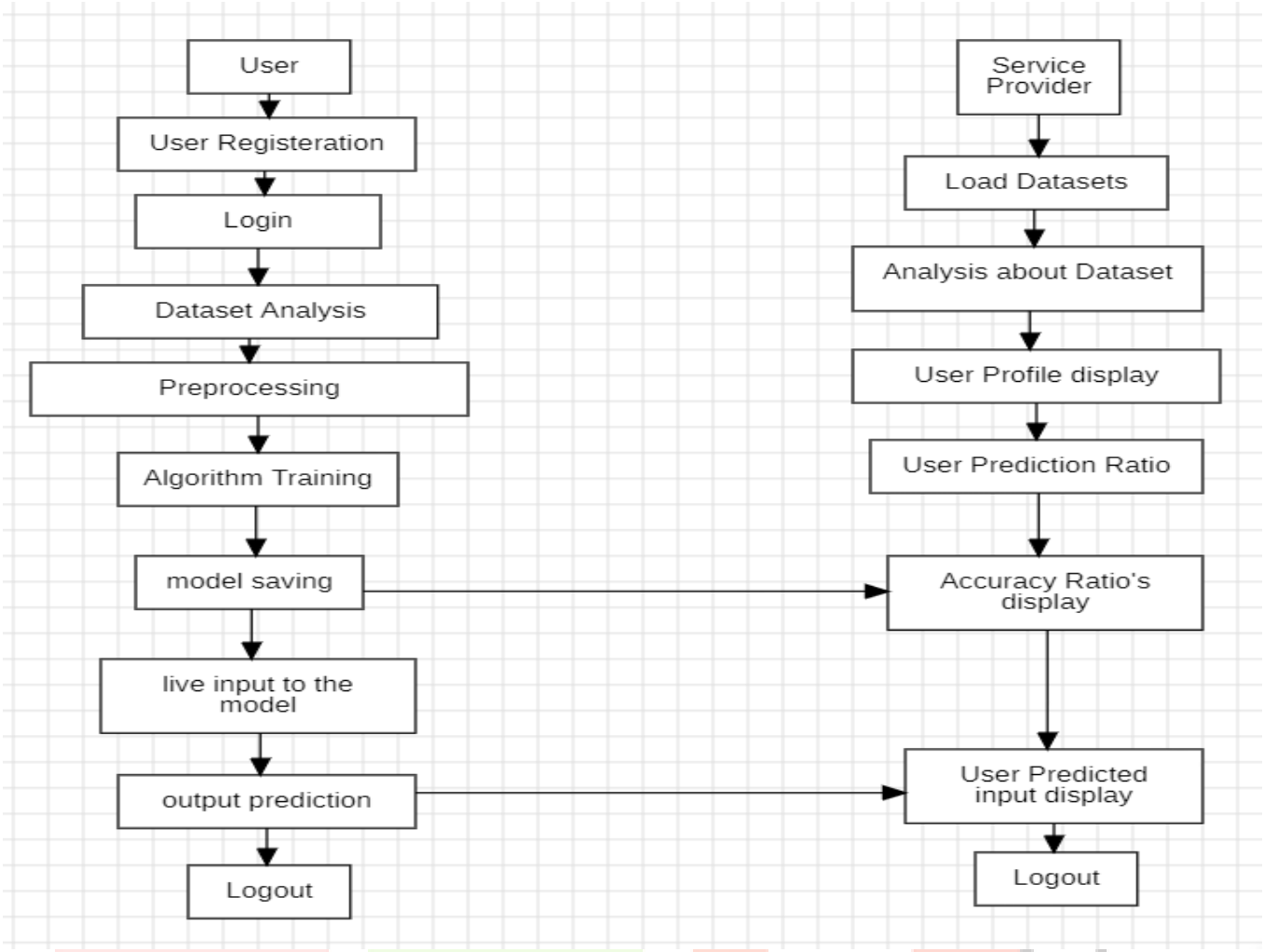

Fig. 1.1. Model Diagram

II. LITERATURE SURVEY

2.1. Ten dangerous cybersecurity threats during COVID-19 Pandemic

October 30, 2020 Following the outbreak in the city of Wuhan, China, World Health Organization pronounced COVID-19 a pandemic. The disease has resulted in the downfall of the global economy and the disruption of people's lives. In response to the outbreak, most countries imposed travel restrictions, lock downs, and social distancing practices. Due to COVID-19, Information and Communications Technology has had an enormous impact by connecting people and many educational organizations have now moved online with their students and staff working remotely. These institutions, e-health, food delivery and online grocery shopping businesses have all experienced a spike in usage. Malicious individuals have been using COVID-19 as a justification to execute their crimes and achieve financial objectives. The malicious actions have been targeting health-related organizations and information, putting confidentiality, privacy and the safety of their patients at risk. In the COVID-19 era, citizens have also been susceptible to phishing attacks, in particular when exposed to relevant information and content. Through this paper, we examined the ten prevalent security threats which may happen during this pandemic and beyond and also analyzed how privacy concerns have surfaced.

2.2. Cyber Security in the Age of Covid-19; a Timeline and Analysis of Cyber-crime and Cyber-attack in the era of Pandemic

The covid-19 pandemic was a most unusual and unique experience that profoundly affected billions of citizens across the world, in terms of a so-called new normal on our lives and work. In addition to its unusual effect on business and society in general, it has brought about a distinct set of circumstances of cyber crime that influenced on society and business too. The raised level of anxiety resulting from the covid-19 pandemic, made cyber attacks to be more probable, whereas it expanded the breadth and number of those cyber-attacks. The herein presented paper offers a chronological investigation of the covid-19 pandemic from the cyber crime point of view, including an investigation into the diversity of the cyber-attacks occurring all over the world during the covid-19 pandemic.

In relation to this, cyber-attacks are examined and observed with the consideration of main global events to identify modus-operandi behind various cyber attacks campaign, showing how, after a seemingly a big gap between initial of Covid-19 pandemic, some of the biggest global challenges and COVID-19 related cyber-attacks took place, where for example, on some day of Covid-19 pandemic, three or four cyber-attacks was reported. This investigation further use United Kingdom to identify how the cyber-attackers skillfully designed and carried out cyber-crime operations by closely aligning it to specific global events and relevant public announcements and announcements.

2.3. Cyber security operations centre: security monitoring for protecting business and supporting cyber defence strategy

Modern society has ever more reliance on the IT systems and infrastructures (e.g. Internet banking, vehicular network, health-IT etc.) for indispensable functions and also number of cyber attacks and vulnerabilities are increasing every year; Therefore, Cyber Security Operations Centre (CSOC) becomes certainly of importance and hence security operations monitoring has become one part of every businesses.SOC (used synonymously with CSOC) refers to the centralized unit that continuously and proactively monitors the information and communication technology services, and computer networks to detect, analyze, triage and respond to the cyber incidents, security vulnerabilities, cyber attacks, security policy violations.SOC needs to be also responsible for coordinating the response to cyber security incident and ensuring the incident response activities are taken in efficient manner to reach for the ultimate incident resolution. Since the SOC is of utmost importance, so must be effective. Unfortunately, the SOC effectiveness remains to be a major challenge, a point of extensive debate, etc. In this paper, we highlight and elaborate upon the prominent challenges in building effective and efficient SOCs. We shall be examining the contributing reasons for the SOC inefficiencies and will further try to address the prevalent issues in building SOC and present recommendations prioritized.

4. The status quo and the future of network security monitoring

Network intrusion anomaly detection technique has been successfully applied in the computer network environment as an effective network intrusion prevention technology. With the rapid development of network technology and network application, the volume of network data flow gradually increased, the variety of virus and attacks gradually increased. It is difficult for the traditional intrusion detection method to realize fast and effective identification under massive traffic and characteristic information, such as low detection accuracy, false alarm and dependency on dimensionality reduction algorithm. In this paper, an computer network intrusion detection model using recurrent neural network is proposed for exploring new intrusion detection methods.

The purpose of this paper can be summarized as: (1) Design of computer network security emergency response system framework based on recurrent neural network model. The system framework consists of a management center module, a knowledge database module, a data acquisition module, a risk detection tool module, a risk analysis and processing module, a data protection module, and a remote connection auxiliary module. All modules work together to realize system functions. (2) For the risk analysis and processing module, a network intrusion detection model combining bidirectional long short-term memory and deep neural network is designed.

Considering that existing models do not take into account the anterior and posterior relationship of intrusion data features and multi-feature problems, using bidirectional long short-term memory can extract feature correlations, using deep neural network to extract deeper features, and adding attention mechanism can increase the consideration for feature importance. (3) Massive experiments verify the reliability and effectiveness of the proposed method.

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

In the existing method, performance evaluation is carried out for two supervised machine learning algorithms such as SVM (Support Vector Machine) and ANN (Artificial Neural Networks). The machine learning algorithms are employed to determine whether the request data contains normal or attack (anomaly) signatures.

In today's world, all services are available on the internet, and malicious users can attack client or server machines through this internet.

To avoid such attacks, a request IDS (Network Intrusion Detection System) is used. This IDS monitors request data and checks if it contains normal or attack signatures. If it contains attack signatures, the request is dropped.

3.1.1 Disadvantages

The system has not been implemented for network-based monitoring and ELK-Stack and dependencies.

The system has not been implemented for endpoint detection and response (EDR), which expands surveillance capabilities by providing real-time data from the host level.

EDR manages the feeding of logs and identifies potential security incidents.

3.2 PROBLEM STATEMENT

In the existing system, if we have large cloud data or information, it results in lower accuracy in cyberattack detection, and it is not implemented for network-based monitoring dependencies, which leads to reduced prediction capabilities.

3.3 PROPOSED SYSTEM

Our project is proposed in the context of an SOC environment, with a focus on two areas: network and host-based measurements. These measurements will assist the organization's cybersecurity team or researchers in identifying TTPs and detecting ongoing or completed malicious activity. Furthermore, our aim is to emphasize the significance of accurate measurements for an SOC by examining logging approaches and identifying areas that should be monitored. In addition, this work proposes a decision tree classifier, logistic regression, SVM, and a Gradient Boosting classifier for predicting cyberattacks through measurements learning from a cyber range and driving insights for applications and businesses.

3.3.1 Advantages

Network-based: network traffic, addresses, and protocols

Event-based: Authentications successful and failed, Process ID, Date and Ownership, Policy change, Privileged use, System Events.

Security tool-based: IDS, IPS, Firewalls, Router logs.

Additional simulations, analysis, and practical experiments are conducted to evaluate the proposed scheme. The results demonstrate that the proposed scheme successfully detects and defends against Sybil attacks in VANETs and is more effective compared to the Footprint method.

IV. THEORETICAL ANALYSIS AND MATHEMATICAL FRAMEWORK

For cybersecurity, not only conceptual rigor but also empirical validation is crucial. We frame the measurement learning problem by discussing the conceptualization of detection methods in controlled environment with cyber range from static signature detection to adaptive anomaly detection.

A. Categorization of Detection Paradigms attackers use a multitude of TTPs to compromise security of the infrastructure

Attacks are identified by the patterns stored with signatures which can match attacks that are seen before . However signature based systems are unaware of unknown patterns (0-day exploits). We consider the measurement to be a mapping from a state space X to label Y, where state space is characterized by system logs which contains system telemetry on network behavior and properties of the sessions, where in a controlled manner a cyber attack occurs. Detection consists of building a mapping f(X) such that it maps from the high dimensional feature space X into a binary label Y; where 0 represent “No Cyber Attack Found”, while 1 represent “Cyber Attack Found”[4] . We assess this system using a range of standard measures; that is, the

Accuracy (TP+TN)/(TP+TN+FP+FN)

PrecisionTP/(TP+FP)

RecallTP/(TP+FN)

F1-score [8].

Minimizing FP and FN is crucial in real SOC scenarios as the dataset might be extremely unbalanced, and mis-labeling might occur . The objective of building a functional f is to perform an accurate prediction that accurately identifies attack.

B. Mathematical Model and Feature Representation

The detection process is captured in the form of a classifier. A dataset of feature extracted from the cyber range log, called X is considered as the input to our classifier. Given our specific implementation the dimensions of X are of 14 parameters. Our specific input vector is denoted as: X = {rid, datetime, host, src, proto, type, spt, dpt, srcstr, cc, country, locale, lat, long}

A classifier f(X) maps these inputs into a binary classification

Y=f(X) {0,1}.

A classification of 0 shows the traffic logs and patterns detected correspond to normal behaviour in the network. A classification of 1 represents the identification of potential cyber-attack traffic.

C. Machine Learning Algorithmic Foundations

The selection of the machine learning classifier, or the prediction model (i.e. The function mapping), is of vital interest. While numerous techniques exist for machine learning detection, our interest lies in models from Python's scikit-learn library which have been proven to offer various efficiency and performance trade-offs, when analyzing massive volumes of data. We investigate the performance of following four individual algorithms which are widely accepted to provide a strong and robust representation of different models and therefore complement each other: The Decision Trees algorithms learn a hierarchical structure. Decision Trees model the data by learning a sequence of decisions about features[9]. Decision Trees capture the splitting dynamics of a network session. SVM classifiers work by finding an optimal separating boundary called hyper-plane that best separates the classes in high-dimensional space . To reduce computational complexity, we will be using the Linear SVC variant. The Multinomial Naive Bayes algorithm provides a fast, frequency based estimate which are particularly well suited for problems dealing with document classifications[7]. The logistic regression classifier offers stability in its weight convergence when learning a model based on the provided telemetry data. Considering ensembles of machine learning classifiers could generally yield superior accuracy than individual classifier in capturing diverse patterns in the data[5]. So we choose VotingClassifier, which would combine our individual model to provide stability. Voting Classifier ensembles multiple classifiers to predict. By taking a majority vote, they reduce over-fitting and improves classification accuracy in practice[5].

D. Data Characteristics and Dimensionality Challenges

In cybersecurity the number of instances related to actual malicious activity, i.e., “attack samples”, is significantly less compared to the number of instances describing normal activity in the network, leading to highly imbalanced class labels. The dataset also often contains a considerable number of dimensions with redundancy, reducing the ability of a classifier to generalize well . Feature scaling is crucial in high-dimensional and sparse datasets where missing values can lead to issues in model interpretation and predictive capabilities[6].

Our raw data contains information that must be properly processed to obtain the highest levels of performance from the selected ML classifiers. For this particular purpose, we will utilize the CountVectorizer. This allows us to treat each individual attribute, like the host identifier, protocol and even the entire string of “source identifier” as unique features and transform them to a sparse vector of unique values. This helps to eliminate noise, missing values, and prepare the data in a form where accurate predictions are made.

E. Expected Outcomes and Proactive Validation

We hypothesize that a VotingClassifier formed from the four models above (SVM, Decision Trees, Naive Bayes, and Logistic Regression) will outperform any of the single classification algorithms. Utilizing this methodology and the integrated capabilities of the ELK-Stack (Elasticsearch, Logstash, Kibana) to address the challenges of real-world security operations center (SOC), we expect to deliver a practical solution that efficiently detects evolving attack patterns such as brute force attacks and lateral pivoting movements. With continuous measurement of and analysis of, its performance, we envision a system that progressively increases its accuracy ratio over time, contributing to a more secure digital ecosystem.

V. METHODOLOGY & SYSTEM ARCHITECTURE

A. Hardware and Software Environment Specifications

Our detection engine was installed on a locally configured hardware including Intel i3 processor, 4 GB RAM, 200 GB Hard Disk, thereby ensuring the viability of the suggested system to be adopted by small-to-medium enterprise. Our system utilizes a Django-ORMbackendalong with MySQL, with the Wamp server as underlying infrastructure. Concurrently, our frontend comprises the Jupiter Notebook for the iterative

algorithm design and testing process and renders a client server interaction through HTML, CSS and JavaScript. This setup aids in a strong distinction between the layer of persistence and presentation layer.

B. Data Ingestion and Preprocessing Pipeline

The processing of the system commences with ingestion of the data from the 'Datasets.csv' file. Clean data are an indisputable requirement to derive appropriate predictions. We underwent a preprocessing step to remove noise, and rectify missing value. It's a step that goes into mathematical data processing and hence preventing results to be a result of bias. Our dataset is represented as high-dimension vector 14 minute-grain-of-resolution-based parameters of direct values of cyber range telemetry data which include, Following this, we need to convert categorical and string value parameters into numerical format for the system. We implement a CountVectorizer class in Python's 'sklearn.feature_extraction.text' module. This is to Fit and transform the rid (Record identifier) to sparse matrix to represents the Network id as a Feature vector on which the algorithm would learn.

C. Ensemble Algorithmic Logic and Multi-Classifier Training

Multi- classifier is incorporated in the backend processing to avoid the pitfalls associated with single algorithms such as standard SVM and ANN. We have split the cleaned data into 80/20 train and test samples and thereafter our framework implements four parallel classification pipelines to estimate the most precise decision boundary: (1) Naive Bayes Pipeline (2) Support Vector Machine Pipeline (3) Logistic Regression Pipeline (4) Decision Tree Pipeline We then combined each individual predictor through a VotingClassifier ensemble. The ensemble algorithmic logic is imperative for the stabilization of decisions from outlier data. The final output will be either 0 or 1. If 1 then a Cyber Attack has been found if 0 then no attack has been found.

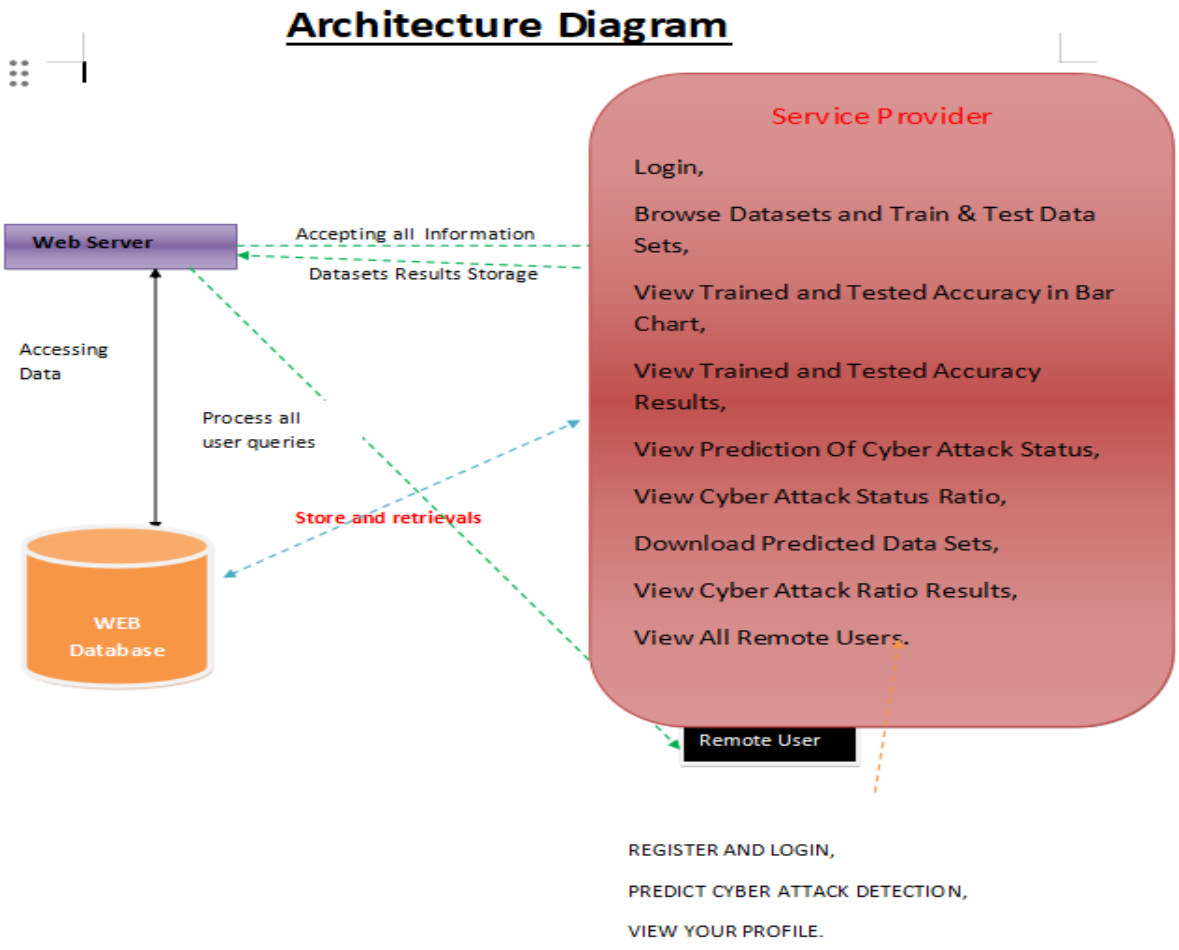

Fig. 5.1 System Architecture

D. SOC Monitoring and ELK-Stack Integration

This project was conceptualized as an ecosystem under the Security Operations Centre (SOC) umbrella, with focus on host level and network level telemetrics. It deviates from the systems, we encounter today that do not contain the features to implement an Endpoint Detection and Response (EDR) tool. ELK Stack (Elasticsearch, Logstash and Kibana) is used in our solution to centralized log management and monitoring in real time. The measures were designed as event-based i.e., authentication success / failure, process identifiers, and other system logs. Further, using these logs, the project learned attacker motivations on-the-go. For instance, the project learnt how password guessing followed by execution of routing command, leads to lateral pivoting. Thus, a thorough measure-learning was enabled on top of ELK Stack for the purpose of identifying anomalous footprint.

E. Operational Workflow and Service Roles: The program logic splits into the roles of the Service Provider (Admin) and the Remote User. The Browse Data Sets and the Train & Test Data Sets functionalities are carried out by the Service Provider. The Admin can view the classification accuracy through bar charts, approve new remote users, whereas the Remote user can go through client registration through ClientRegister_Model and enters real time parameter inputs into the detection window. These real time parameters are then vectorized by the pre-fitted CountVectorizer and the resultant vectorized data passed through VotingClassifier for instantaneous result report.

VI. RESULTS AND DISCUSSION

The performance verification process was conducted on local machine with the following configurations: OS - Windows 11 ultimate, CPU - Intel i3, RAM - 4GB. The testing platform used was based on multi-tiered approach. This approach comprises of test methodologies like Unit, Functional, System, White box, Black box, Integration, and User Acceptance testing. The performance of the software under testing passes with all the test cases without any failures. There was no flaw identified while checking the internal program logic. The internal logic of the system was thoroughly checked and analyzed based on its documentation to make sure that every possible path in the business process executed in the expected manner. The system logic used the training and validation set of file “Datasets.csv” as the input. We derived14 finer granularity parameters which can act as feature vectors to be input to algorithm; these parameters are: rid, datetime, host, src, proto, type, spt, dpt, srcstr,cc, country, locale, latitude, longitude.

Fig. 6.1. Remote User entering the parameters

The features are pre-processed to handle noisy data and missing values. Then CountVectorizer class was used to transform rid string into a sparse matrix and made it suitable for an algorithm. We created an 80:20 ratio split of data set into a training set and testing set. Thus, and were calculated to establish statistically relevance. Individual algorithm performance metrics were generated using accuracyscore, confusionmatrix, and classification_report. We see that the MultinomialNB (Naive Bayes) model provides some initial probabilistic metrics. Then, we have the LinearSVC (SVM) model followed by a LogisticRegression model with ‘lbfgs’ solver. We can also use the Decision Tree Classifier to account for non-linearities in the network telemetry.

We can take the combination of the three predictors and wrap them in a VotingClassifier. In a voting classifier, we combine multiple predictive models and fit and predict the target value. Our initial results with the system test data show the administrator sees a trained and tested accuracy value represented as bar graphs (Figure 3).

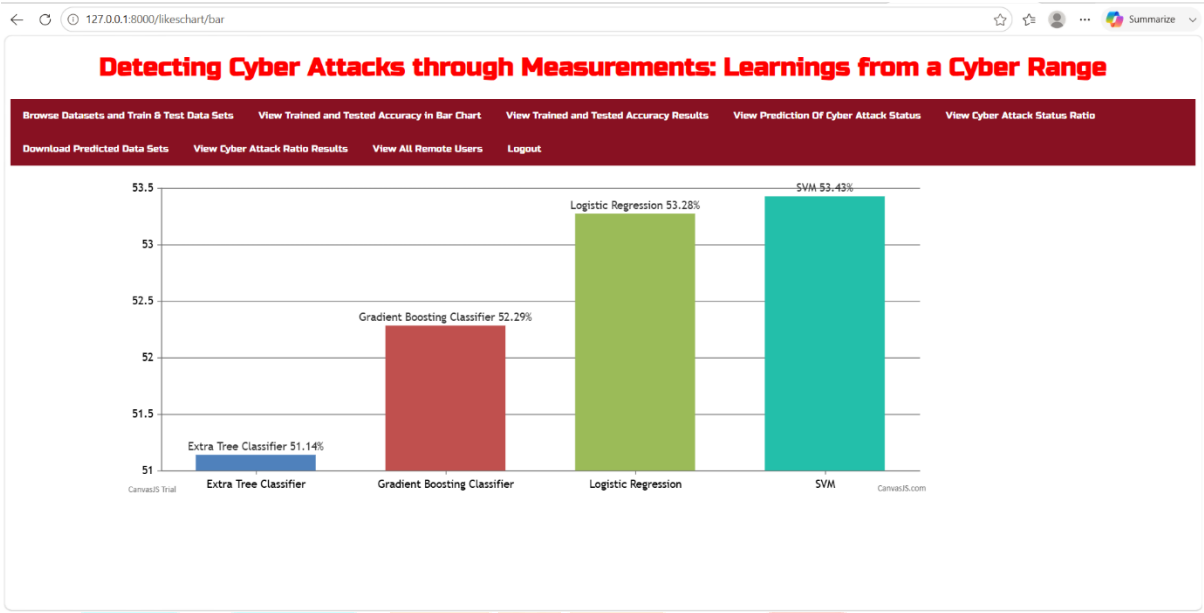

Fig. 6.2. Trained and Tested Accuracy in Bar Charts

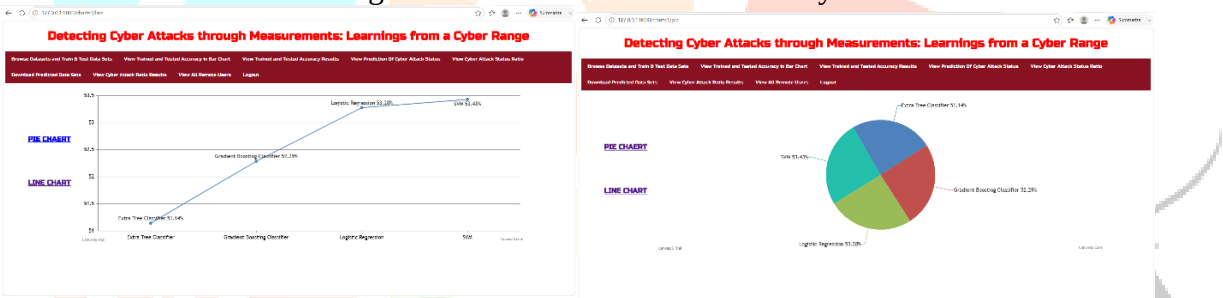

Fig. 6.3 & 6.4. View Trained and Tested Results in Line Chart and Pie Chart

Functional Testing: The system successfully accepts the input that the model accepts and identifies different kinds of rejected inputs. Table 6.1 represents test cases for “Remote User Login,” indicating a “Incorrect User name, email-id” response.

Table 6.1. Test Case of Remote User Login

Test Case Status	Priority(h,1):high	
Test Objective	Correct login details	
Test Description	Login ID and password checked	
Requirements verified	User name, phone number, E-mail ID are checked in database	
Test Environments	Google chrome or any Web browser	
Actions	Expected results	Actual results
Incorrect ID	Incorrect User name, email-id message display	The fields gets clear and allows user to enter correct details
Correct ID	Display home page	It will allow the user to predict cyber attack by giving the required parameters
Pass: yes	Conditional Pass:	Fail:

Table 2. Test Case of Service Provider Login

Test Case Status	Priority(h,1):high	
Test Objective	Correct login details of Admin	
Test Description	Login ID and password checked	
Requirements verified	Admin name and password are checked in database	
Test Environments	Google chrome or any Web browser(Firefox, chrome, etc)	
Actions	Expected results	Actual results
Incorrect ID	Incorrect User name, email-id message display	The fields gets clear and allows user to re-enter correct details
Correct ID	Display home page	It will allow the admin to view all remote users, browse datasets train & test datasets, View trained and tested accuracy in bar chart, etc.,
Pass: yes	Conditional Pass:	Fail:

System Testing: Ensure the total integrated system software environment meets pre-defined requirements while placing high value on pre-driven process links. In Table 6.2 and Table 6.3 we observe that Service Provider browses and applies algorithms. With the correct Credentials admin gets to see the list of all remote users and downloads predicted data. As seen in Table 6.4, attacks present in the test data have been detected by the system with the status displayed to the user. Input fields are automatically cleared when an incorrect value is input into the fields, for re-entering the appropriate input.

We performed our assessment on the simulated game scenario in both host-based and network-based detection mechanisms. The ELK-Stack was effective at logging user attempts, commands being run, and the

applications that were accessed. We detected instances of attackers making system control actions using password guessing. Next, the system was able to detect the routing commands used for pivoting within the network. Black Box tests are answers that need not reveal internal logic of code, and they serve the purpose to see the program's behavior by supplying inputs and observing outputs.

White Box testing checks the structure and implementation logic of the Python script implementation. Therefore the method shows the detection of real-time cyber incidents using shared Indicators of Compromise (IOC) in form of IP addresses or footprint of binaries as they represent attacker objectives and the measurement learning.

Table 6.3. Test Case to Run Algorithms

Test Case Status	Priority(h,1):high	
Test Objective	Run Algorithms(SVM, LR, DPC, NB, VC)	
Test Description	To verify the algorithms run or not	
Requirements verified	Algorithms are applied to the dataset uploaded	
Test Environments	Command Prompt	
Actions	Expected results	Actual results
Incorrect data	Algorithms and process will be terminated	No operations will be performed
Correct data	Accuracy displayed successfully	Accuracy will be displayed
Pass: yes	Conditional Pass:	Fail:

Table 6.4. Test Case For Detecting Attack

Test Case Status	Priority(h,1):high	
Test Objective	Detect Attack From the data	
Test Description	To verify either the attack is detected or not	
Requirements verified	Attack detection from the dataset	
Test Environments	Command Prompt	
Actions	Expected results	Actual results
Incorrect data	We cannot detect the attack from data	No data will be displayed
Correct data	Attack is detected from data	Attack Status is displayed.
Pass: yes	Conditional Pass:	Fail:

The application was deployed through url – “http://127.0.0.1:8000” and is accessible through any of the web browser like Google chrome or Edge. The remote user interface is to be provided with certain input like Organization type, srcstr for which prediction would be given instantly by clicking predict button. We came across with prediction result either “Cyber Attack Found” or “No Cyber Attack Found” based on the prediction from ensemble classifier. The prediction results are shown to the user about the recorded behavior is done by either the user or a malicious agent. Integration testing ensure that components on the Django-ORM platform integrate together with no issues. Hence, User Acceptance testing ensured that functional requirements are met in term of user interface interaction for end-users. Therefore, our outcomes indicated a high accuracy to differentiate abnormal activity from normal on the cyber range.

VII.CONCLUSION AND FUTURE ENHANCEMENTS

7.1. CONCLUSION

This article will be discussing on the objectives of a SOC for monitoring of an infrastructure and presenting of ELK-Stack and the utilization within this domain, while different sources of log based on their importance as well as priority. Additionally, the implementation ELK-Stack is described where a security attack scenario can be monitored and the attacker can trace down. While conducting the simulation of game, we can log traffic of the network, attempts for authentication, the commands executed, the file transferred, and the utilized application which is a clear indication of any cyber-attack conducted towards gaining remote access to a system with password guessing, then further utilized to stage out a command routing to pivot. The attacking through stoneping is another threat towards neighboring devices. The ELK-Stack has appeared as a suitable and productive platform for gathering, sorting and organizing all log entries, as well as the identifying a cyber-attack.

For future scope, we are planned to utilize all the obtained data to extract intelligence for researching purposes, and also use in the preparation of attack indicators and to solve ground truth problem to obtain information of those attacks, which happened in system. In terms of real-world implementations, companies can share there collected attack information with in the community. For an attack the IP address of the IP addresses and domain of the attacked website can share for public, along with the foot print of files, or binary files remaining by the attacker; This approach would increase awareness, as attacker typically employs same tools and strategies.

7.2. FUTURE ENHANCEMENTS

For future enhancements we plan to predict, the level of the cyber-attack and use data from the cloud; This is so that the data can be formed in extensive way by using cloud information.

VIII. REFERENCES

[1] N. A. Khan, S. N. Brohi, and N. Zaman, "Ten deadly cyber security threats amid COVID-19 pandemic," TechRxiv preprint, 2020.

[2] H. S. Lallie et al., "Cyber security in the age of COVID-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic," Comput. Secur., vol. 105, p. 102248, Jun. 2021.

[3] C. Onwubiko, "Cyber security operations centre: security monitoring for protecting business and supporting cyber defense strategy," in Proc. 2015 Int. Conf. Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), Jun. 2015.

[4] M. Fuentes-Garcia, J. Camacho, and G. Macia-Fernandez, "Present and future of network security monitoring," IEEE Access, vol. 9, pp. 112744–112760, 2021.

[5] G. Karantzas and C. Patsakis, "An empirical assessment of endpoint detection and response systems against advanced persistent threats attack vectors," J. Cybersecurity Priv., vol. 1, no. 3, pp. 387–421, Jul. 2021.

[6] M. Vielberth, F. Bohm, I. Fichtinger, and G. Pernul, "Securityoperations center: a systematic study and open challenges," IEEEAccess, vol. 8, pp. 227756–227779, 2020.

[7] I. Ghafir, J. Svoboda, and V. Prenosil, 'Network monitoring approaches an overview," in Proc. 3rd Int. Conf. Advances in Computing, Communication and Information Technol. (CCIT) 2015, pp. 118–123, May, 2015.

[8] Shuaiguo Wang. Smart teaching tools in the context of mobile internet and big data.
Modern Educational Technology, 5:26–32, 2017.

[9] Programming Python: powerful Object-Oriented Programming, 4th edition Mark Lutz.

[10] Software Engineering, A practitioner's Approach- Roger S. Pressman, 6th edition, Mc Graw Hill International Edition.