A

Major Project

On

Detecting Cyber Attacks through Measurements: Learning from a Cyber Range

(Submitted in partial fulfillment of the requirements for the award of the Degree)

BACHELOR IN TECHNOLOGY

In

INFORMATION TECHNOLOGY

By

Muddam Ranadeep

(207R1A1295)

Under the Guidance of

Dr.V.Malsoru

(Professor)

DEPARTMENT OF INFORMATION TECHNOLOGY

CMR TECHNICAL CAMPUS

UGC AUTONOMOUS

(Accredited by NAAC, NBA, Permanently Affiliated to JNTUH, Approved by AICTE,
NewDelhi) Recognized Under Section 2(f) & 12(B) of the UGCAct.1956,
Kandlakoya (V), Medchal Road, Hyderabad-501401

2020-2024

DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled **Detecting Cyber Attacks through Measurements: Learning from a Cyber Range** submitted by **Muddam Ranadeep (207R1A1295)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Information Technology from the Jawaharlal Nehru Technological University Hyderabad, is a record of the Bonafide work carried out by him under our guidance and supervision during the academic year 2023-2024

The results embodied in this project has not been submitted in any other University or institute for the award of any degree or diploma.

Dr.V.Malsoru
(Professor)
INTERNAL GUIDE

Dr. A. Raji Reddy
DIRECTOR

Dr. B. Kavitha Rani
HEAD OF THE DEPARTMENT

EXTERNAL EXAMINAR

Submitted for Viva voice Examination held on______________________________

ACKNOWLEDGEMENT

Apart from the efforts, the success of any project depends largely on the encouragement and guidelines of many others. I take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

I take this opportunity to express our profound gratitude and deep regard to our guide faculty, **Dr.V.Malsoru**, Professor, for his exemplary guidance, monitoring and constant encouragement throughout the course of this Subject presentation.

I also take this opportunity to express our profound gratitude to our Project Review Committee (PRC) Coordinator, **Mr.MD Sajid Pasha** Associate Professor. The blessing, help and guidance given by him shall carry us a long way in the journey of life on which I are about to embark.

I also thankful to the Head of the Department **Dr. B. Kavitha Rani** for providing excellent infrastructure and a nice atmosphere for completing this project successfully.

I would like to express our sincere gratitude to **Dr. M. Ahmed Ali Baig,** Dean Administration, **Dr. D T V Dharmajee Rao,** Dean Academics, **Dr. Ashutosh Saxena,** Dean R&D for encouragement throughout the course of this presentation.

I obliged to our Director **Dr. A. Raji Reddy** for being cooperative throughout the course of this presentation.

I wish to express our sincere gratitude to the Management of **CMR Technical Campus**, Hyderabad, **Sri.Ch. Gopal Reddy,** Honourable Chairman, S**mt.C. Vasantha Latha,** Honourable secretary, **Sri.C. Abhinav Reddy,** Honourable Chief Executive Officer.

The guidance and support received from all the members of **CMR TECHNICAL CAMPUS** who contributed and who are contributing to this project, was vital for the success of the project. I grate full for their constant support and help.

Finally, I would like to take this opportunity thank our family for their constant encouragement without which this assignment would not be possible. I sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project.

Muddam Ranadeep

(207R1A1295)

ABSTRACT

Nowadays, it is hard to see an organization without a digital presence, while our modern society relies on a wide range of activities like banking, government services, commerce, or education that are offered online. Even more, the recent years have pushed the limits of digital transformation for multiple organizations, companies, and educational institutions. This transition occurred without any prior planning or preparation and at an unprecedented scale . As the globe is converging towards a technology driven society, cyber attacks and cyber crime campaigns are blooming. Recent reports show that cyber crime is growing in severity and frequency, competing with the traditional crime in both the number of incidents and revenue.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, different type of algorithms is trained to make classifications or predictions, and to uncover key insights in this project. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics.

Machine learning algorithms build a model based on this project data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of datasets, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

LIST OF FIGURES

LIST OF SCREENSHOTS

CONTENTS

TITLE **PAGE NO**

Detecting Cyber Attacks through Measurements: Learnings from a Cyber Range

1. INTRODUCTION

1. INTRODUCTION

1.1 INTRODUCTION

Detecting cyber attacks through measurements, as demonstrated in Cyber Range projects,involves employing data-driven methods to identify and mitigate threats. These projects simulate realistic attack scenarios, providing valuable data for developing and testing detection algorithms.By analyzing network traffic and system behavior, researchers can detect anomalies and patterns indicative of cyber attacks. This approach enables the development of more effective and efficient cyber security measures.

1.2 PROBLEM STATEMENT

The project aims to develop a system for detecting cyber attacks by analyzing measurement data from network traffic and system logs in a simulated Cyber Range environment. The system will identify attack patterns, detect anomalies in real-time, and provide alerts to security analysts. It will also incorporate machine learning to improve accuracy. Additionally, the project includes creating a training program to educate cybersecurity professionals on using the system effectively.

1.3 OBJECTIVES

The objectives of the project are to develop a system for detecting cyber attacks using measurement data within a Cyber Range environment. This system will identify attack patterns, analyze network traffic and system logs, generate alerts for potential attacks, and incorporate machine learning for improved accuracy. Additionally, the project aims to create a training program to educate cybersecurity professionals on using the detection system effectively in a simulated setting.

1.4 LIMITATIONS

The limitations of the project include the challenge of accurately distinguishing between normal network behavior and malicious activity, which can lead to false positives or false negatives in attack detection. The availability and quality of measurement data may also impact the effectiveness of the detection system. Additionally, the complexity of cyber attacks and the constantly evolving nature of threats may require regular updates and adjustments to the system to maintain its effectiveness.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

2.1 INTRODUCTION

Nowadays, it is hard to see an organization without a digital presence, while our modern society relies on a wide range of activities like banking, government services, commerce, or education that are offered online. Even more, the recent years have pushed the limits of digital transformation for multiple organizations, companies, and educational institutions. This transition occurred without any prior planning or preparation and at an unprecedented scale . As the globe is converging towards a technology driven society, cyber attacks and cyber crime campaigns are blooming. Recent reports show that cyber crime is growing in severity and frequency, competing with the traditional crime in both the number of incidents and revenue

2.2 EXISTING SYSTEM

According to Vielberth et al. , the number of the documented breaches for companies has been increased over the last five years by 65%. The average time to detect an incident was 196 days in 2018 plus another 69 days to contain it, meaning that many attacks stayed under the radar for a long period. Vielbeth etal. acknowledge that possible reasons for this belated discovery include: the failure of overview for devices, systems, applications and networks; uncertainty on which assets to monitor and protect; and lack of knowledge in regards to appropriate tools and how to integrate them. Finally, they suggest that organizations can be overpowered by the technological speed adapted by the cyber criminals and the rapidly growing threat landscape.Creating visibility across network assets can improve the overall company security posture, by reducing the severity and eventually the financial loss of a cyber attack. Decreasing the detection time of an incident, directly implies that attackers have less time to wander around the company's infrastructure for snooping into sensitive data or critical resources. Nonetheless, the detection is not sufficient by itself and it should be combined with the rest of the key items referred to in the NIST Cyber Security Framework. Although SOCs offer multiple benefits there are also some challenges when it comes to their implementation. The survey by Vielberth et al. systematically groups difficulties into Processes, People, Governance and Compliance, and Technology. For instance, processes need to be integrated across the whole organization.Additionally a lack of skilled personnel represents a challenge in recruiting and retaining staff, which can be addressed only by raising an awareness culture. Nonetheless, governance and compliance can be difficult to form without unified standards, which results in impediments to security audits and overall assessments. Lastly, even the fact that technology is vast can create issues in choosing the best solution for a particular use case.An SOC can accomplish the monitoring of the infrastructure’s assets at different layers; network-based monitoring refers to the detection mechanisms placed at the network layer, while endpoint-based monitoring is the collection of the mechanisms at the host layer. The latter type offers a more fine grained visibility of the infrastructure’s state. According to Fuentes-Garcia et al. , a network security monitoring system should provide traceability of the processes of the network and systems under monitoring. However, to achieve this view, the setup should incorporate multiple components, such as those described subsequently

2.3 DISADVANTAGES OF EXISTING SYSTEM

The system is not implemented for Network-based Monitoring and ELK-Stack and Dependencies. The system is not implemented Endpoint detection and response (EDR) which never expands the surveillance capabilities by providing real-time collection of information from the host level. The EDR's role is to manage the feeding of logs and to detect potential security incidents

2.4 PROPOSED SYSTEM

This project presents the monitoring capabilities in the context of an SOC enviroment, focusing on two vantage points, namely, network and host based measurments. These measurements can help the cyber security team of the organization or the researchers both to determine the TTPs and identify ongoing or completed malicious activity. Furthermore, we aim to highlight the importance of accurate measurements for the objectives of an SOC by exemplifying the logging approaches and pinpointing the locations where activity should be monitored. Moreover, this work provides examples of tools that can support the operational requirements of an SOC, with a focus on Elasticsearch, Logstash and Kibana (ELK-Stack). In our research, the ELK-Stack is used for the collection, processing, and correlation of different log sources which are essential for the identification of security incidents. Since it is based on the log analysis, the SOC aims to infer whether an incident took place or is in progress within the monitored infrastructure.Finally, we offer directions of how these data can be further utilized for the purposes of cyber security. We provide an overview of the current techniques and methods for infrastructure monitoring in the context of cyber security, by giving focus on security information event management (SIEM) systems. SIEMs are a set of technologies collaborating to provide a comprehensive view of the infrastructure.The SIEM provides the technical foundation for an SOC to function, engaging many necessary processes for early response to security incidents. By building on the ELK-Stack and its dependent applications, one is able to aggregate network traffic, system events, security-related events, and other metrics.

2.5 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, Routers logs Further simulations, analysis, and practical experiments are conducted to evaluate the proposed scheme and compare it with the Footprint [4], the results indicate that the proposed scheme can successfully detect and defend against Sybil attacks in VANETs and more efficiently compared to the Footprint.

3. SYSTEM STUDY

3. SYSTEM STUDY

3.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibilitystudy of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirementsfor the system is essential.

Three key considerations involved in the feasibility analysis are

- Economical Feasibility
- Technical Feasibility
- Social Feasibility

3.1.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of funds that the company can pour into the research and development ofthe system is limited. The expenditure must be justified. Thus, the developed system is well withinthe budget, and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

3.1.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirementsof the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system

3.1.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

4. SYSTEM REQUIREMENTS

4. SYSTEM REQUIREMENTS

4.1 HARDWARE REQUIREMENTS

System	Intel i3 2.4 GHz.
Hard Disk	40 GB.
Floppy Drive	1.44 Mb.
Ram	512 Mb.

4.2 SOFTWARE REQUIREMENTS

Operating system	Windows 11.
Coding Language	Python 3.8
Designing	Html,css,javascript.
Data Base	MySQL.

5. SYSTEM DESIGN

5. SYSTEM DESIGN

5.1 INTRODUCTION

5.1.1.SYSTEM ARCHITECTURE

The architecture of a Cyber Range for detecting cyber attacks through measurements comprises interconnected components designed to simulate real-world network environments. It includes virtualized infrastructure, such as servers, routers, and firewalls, alongside monitoring and analysis tools. The system collects and analyzes data from network traffic, system logs, and sensor outputs to identify anomalous behavior and potential security threats. Advanced algorithms and machine learning techniques are utilized to correlate and analyze the collected data, enabling the detection of cyber attacks in real-time. The architecture fosters proactive defense strategies, allowing organizations to test and improve their incident response capabilities in a controlled environment. Overall, the Cyber Range architecture serves as a critical tool for cybersecurity training, research, and testing

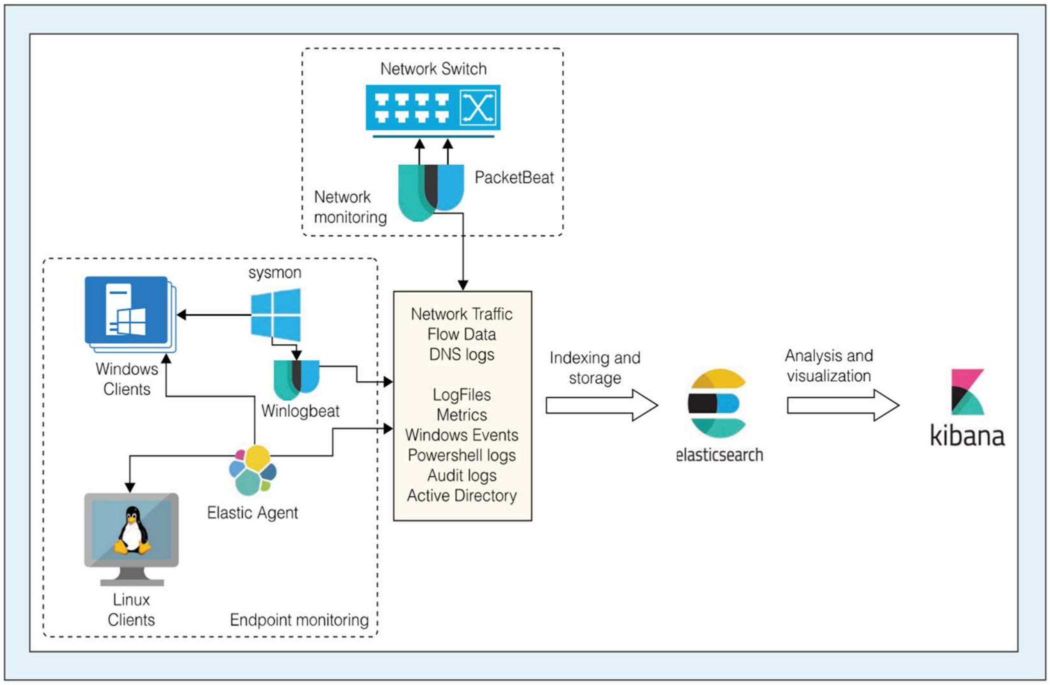


Fig 5.1 System Architecture of Detecting Cyber Attacks

5.2 UNIFIED MODELLING LANGUAGE(UML)

UML represents Unified Modeling Language. UML is an institutionalized universally useful showing dialect in the subject of article situated programming designing. The fashionable is overseen, and become made by way of, the Object Management Group.

The goal is for UML to become a regular dialect for making fashions of item arranged PC programming. In its gift frame UML is contained two noteworthy components: a Meta-show and documentation. Later on, a few type of method or system can also likewise be brought to or related UML.

The Unified Modeling Language is a popular dialect for indicating, Visualization, Constructing and archiving the curios of programming framework, and for business demonstrating and different non-programming frameworks.

The UML speaks to an accumulation of first-rate building practices which have verified fruitful in the showing of full-size and complicated frameworks.

The UML is a essential piece of creating gadgets located programming and the product development method. The UML makes use of commonly graphical documentations to specific the plan of programming venture

5.2.1 GOALS OF UML

The Primary goals in the design of the UML are as follows:

- Provide users a ready-to-use, expressive visual modeling language so that they can develop and exchange meaningful models.
- Provide extendibility and specialization mechanisms to extend the core concepts.
- Be independent of particular programming languages and development process.
- Provide a formal basis for understanding the modeling language.
- Support higher level development concepts such as collaboration, frameworks, patterns andcomponents.
- Integrate best practices.

5.2.2 USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

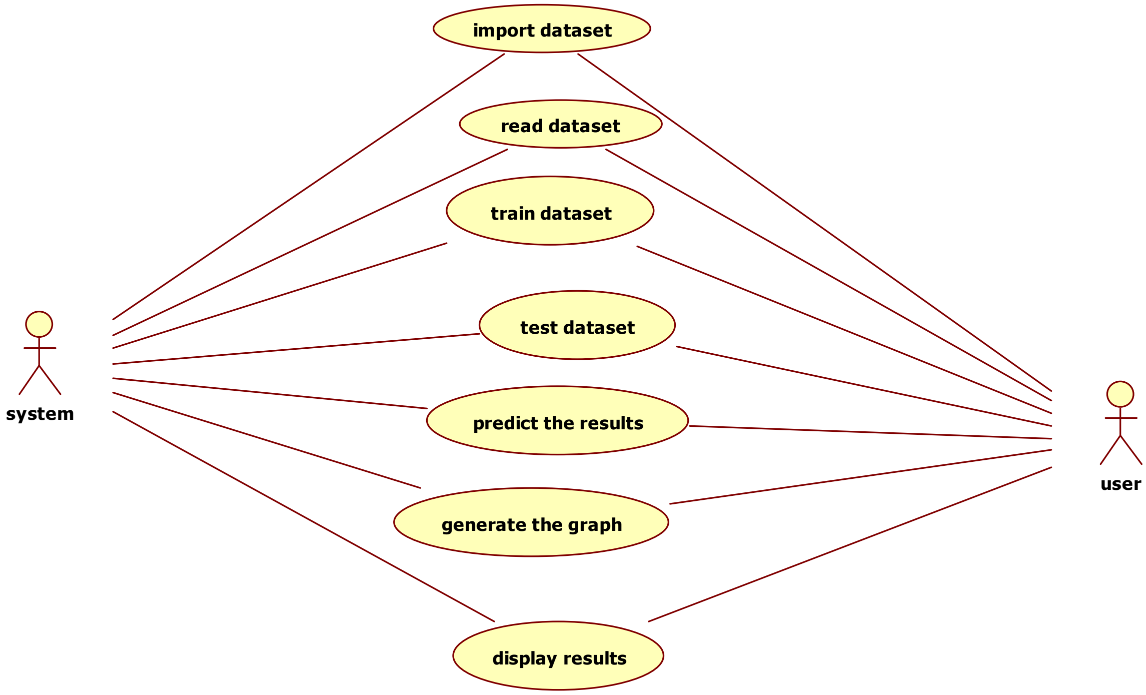


Fig 5.2.2 Use-Case Diagram of Detecting Cyber Attacks

5.2.3 CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

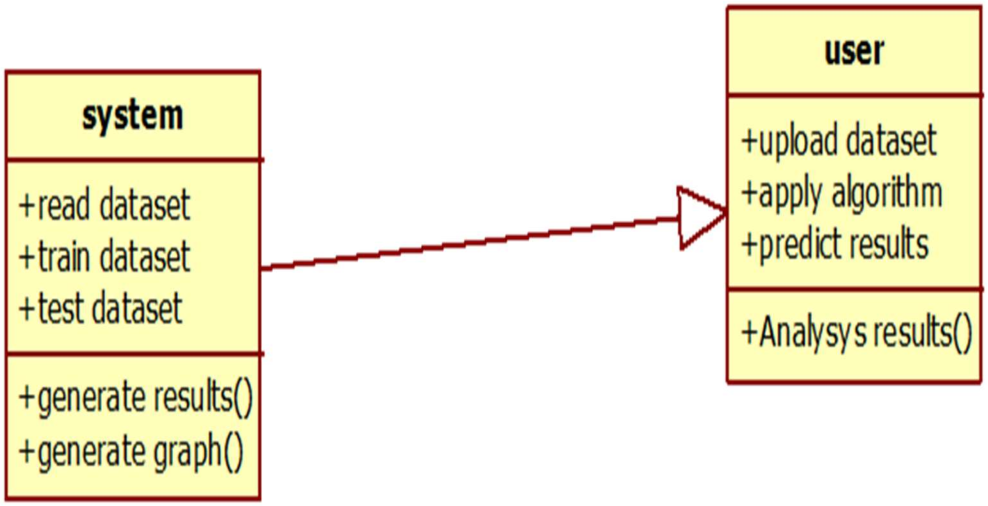


Fig.5.2.3 Class diagram of Detecting Cyber Attacks

5.2.4 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

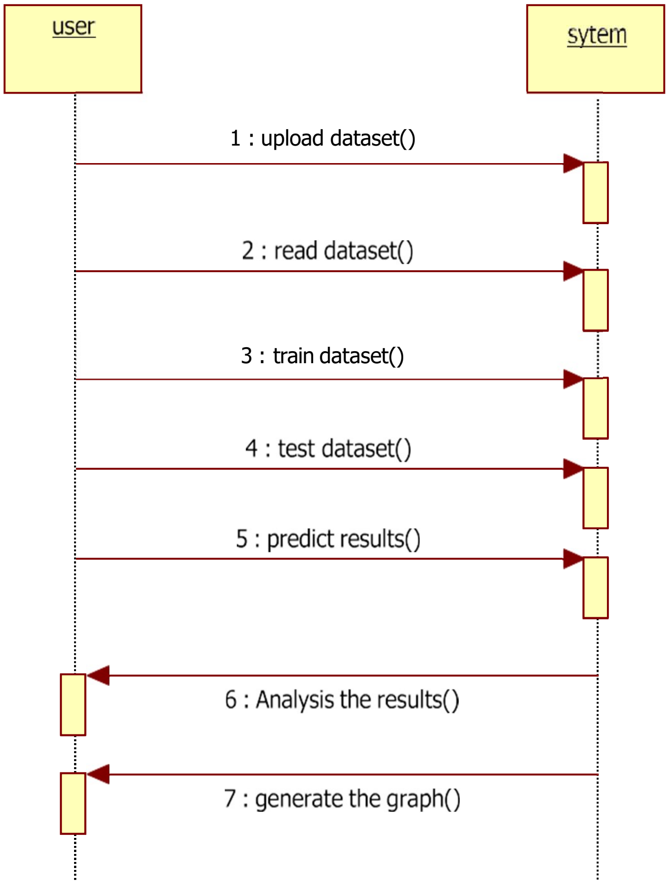


Fig 5.2.4 Sequence diagram of Detecting Cyber Attacks

5.2.5 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

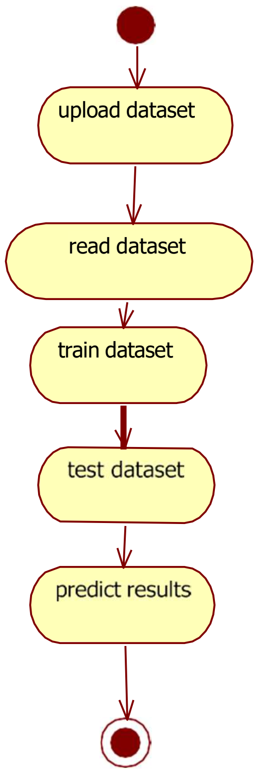


Fig 5.6 Activity diagram of Detecting Cyber Attacks

6. IMPLEMENTATION

6. IMPLEMENTATION

6.1 MODULES

- SERVICE PROVIDER
- VIEW AND AUTHORIZE USERS
- REMOTE USER

6.2 MODULE DESCRIPTION

6.2.1 SERVICE PROVIDER :

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Predicted Type Details, Find Cyber attacks Predicted Type Ratio, Download Predicted Data Sets, View Predicted Cyber attacks Type Ratio Results, View All Remote Users.

6.2.2 VIEW AND AUTHORIZE USERS :

In this module, the admin can view the list of users who all registered.In this, the admin can view the user’s details such as, user name, email, address and admin authorizes the users.

6.2.3 REMOTE USER :

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN,PREDICT Cyber attacks.

6.3 SOURCE CODE

```
from django.db.models import  Count, Avg
from django.shortcuts import render, redirect
from django.db.models import Count
from django.db.models import Q
import datetime
import xlwt
from django.http import HttpResponse
import numpy as np
import nltk
import re
import string
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import CountVectorizer
from nltk.stem.wordnet import WordNetLemmatizer
import pandas as pd
from wordcloud import WordCloud, STOPWORDS
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.metrics import accuracy_score
from sklearn.metrics import f1_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import VotingClassifier
from sklearn.ensemble import RandomForestClassifier
# Create your views here.
from Remote_User.models import ClientRegister_Model,Cyber_model,detection_ratio,detection_accuracy
def serviceproviderlogin(request):
    if request.method  == "POST":
        admin = request.POST.get('username')
        password = request.POST.get('password')
        if admin == "Admin" and password =="Admin":
    detection_accuracy.objects.all().delete()
    return redirect('View_Remote_Users')
```

```
 return render(request,'SProvider/serviceproviderlogin.html')
def View_CyberThreat_Type_Ratio(request):
 detection_ratio.objects.all().delete()
 ratio = ""
 kword = 'Theft'
 print(kword)
 obj = Cyber_model.objects.all().filter(Q(Prediction=kword))
 obj1 = Cyber_model.objects.all()
 count = obj.count();
 count1 = obj1.count();
 ratio = (count / count1) * 100
 if ratio != 0:
     detection_ratio.objects.create(names=kword, ratio=ratio)
 ratio1 = ""
 kword1 = 'Loss'
 print(kword1)
 obj1 = Cyber_model.objects.all().filter(Q(Prediction=kword1))
 obj11 = Cyber_model.objects.all()
 count1 = obj1.count();
 count11 = obj11.count();
 ratio1 = (count1 / count11) * 100
 if ratio1 != 0:
     detection_ratio.objects.create(names=kword1, ratio=ratio1)
 ratio12 = ""
 kword12 = 'Disclosure'
 print(kword12)
 obj12 = Cyber_model.objects.all().filter(Q(Prediction=kword12))
 obj112 = Cyber_model.objects.all()
 count12 = obj12.count();
 count112 = obj112.count();
 ratio12 = (count12 / count112) * 100
 if ratio12 != 0:
     detection_ratio.objects.create(names=kword12, ratio=ratio12)
 ratio123 = ""
 kword123 = 'Hacking'
```

```
    print(kword123)
    obj123 = Cyber_model.objects.all().filter(Q(Prediction=kword123))
    obj1123 = Cyber_model.objects.all()
    count123 = obj123.count();
    count1123 = obj1123.count();
    ratio123 = (count123 / count1123) * 100
    if ratio123 != 0:
        detection_ratio.objects.create(names=kword123, ratio=ratio123)
    ratio123a = ""
    kword123a = 'Improper Disposal'
    print(kword123a)
    obj123a = Cyber_model.objects.all().filter(Q(Prediction=kword123a))
    obj1123a = Cyber_model.objects.all()
    count123a = obj123a.count();
    count1123a = obj1123a.count();
    ratio123a = (count123a / count1123a) * 100
    if ratio123a != 0:
        detection_ratio.objects.create(names=kword123a, ratio=ratio123a)
    obj = detection_ratio.objects.all()
    return render(request, 'SProvider/View_CyberThreat_Type_Ratio.html', {'objs': obj})
    def View_Remote_Users(request):
        obj=ClientRegister_Model.objects.all()
        return render(request,'SProvider/View_Remote_Users.html',{'objects':obj})
def ViewTrendings(request):
    topic = Cyber_model.objects.values('topics').annotate(dcount=Count('topics')).order_by('-dcount')
    return  render(request,'SProvider/ViewTrendings.html',{'objects':topic})
def charts(request,chart_type):
    chart1 = detection_ratio.objects.values('names').annotate(dcount=Avg('ratio'))
    return render(request,"SProvider/charts.html", {'form':chart1, 'chart_type':chart_type})
def charts1(request,chart_type):
    chart1 = detection_accuracy.objects.values('names').annotate(dcount=Avg('ratio'))
    return render(request,"SProvider/charts1.html", {'form':chart1, 'chart_type':chart_type})
def View_Prediction_Of_CyberThreat_Type(request):
    obj =Cyber_model.objects.all()
    return render(request, 'SProvider/View_Prediction_Of_CyberThreat_Type.html', {'list_objects': obj})
```

```
def likeschart(request,like_chart):
    charts =detection_accuracy.objects.values('names').annotate(dcount=Avg('ratio'))
    return render(request,"SProvider/likeschart.html", {'form':charts, 'like_chart':like_chart})
def Download_Trained_DataSets(request):
    response = HttpResponse(content_type='application/ms-excel')
    # decide file name
    response['Content-Disposition'] = 'attachment; filename="Predicted_Data.xls"'
    # creating workbook
    wb = xlwt.Workbook(encoding='utf-8')
    # adding sheet
    ws = wb.add_sheet("sheet1")
    # Sheet header, first row
    row_num = 0
    font_style = xlwt.XFStyle()
    # headers are bold
    font_style.font.bold = True
    # writer = csv.writer(response)
    obj = Cyber_model.objects.all()
    data = obj  # dummy method to fetch data.
    for my_row in data:
        row_num = row_num + 1
        ws.write(row_num, 0, my_row.Name_of_Covered_Entity, font_style)
        ws.write(row_num, 1, my_row.State, font_style)
        ws.write(row_num, 2, my_row.Individuals_Affected, font_style)
        ws.write(row_num, 3, my_row.Date_of_Breach, font_style)
        ws.write(row_num, 4, my_row.Location_of_Breached_Information, font_style)
        ws.write(row_num, 5, my_row.Date_Posted_or_Updated, font_style)
        ws.write(row_num, 6, my_row.breach_start, font_style)
        ws.write(row_num, 7, my_row.year, font_style)
        ws.write(row_num, 8, my_row.Source_Ip, font_style)
        ws.write(row_num, 9, my_row.Destination_Ip, font_style)
        ws.write(row_num, 10, my_row.Prediction, font_style)
      wb.save(response)
      return response
             def train_model(request):
```

```
    detection_accuracy.objects.all().delete()
    data = pd.read_csv("Cyber_Threat.csv")
    # data.replace([np.inf, -np.inf], np.nan, inplace=True)
    mapping = {'Theft': 0,
        'Loss': 1,
        'Disclosure': 2,
        'Hacking': 3,
        'Improper Disposal': 4

        }
    data['Label'] = data['Type_of_Breach'].map(mapping)

    x = data['Name_of_Covered_Entity']
    y = data['Label']
# data.drop(['Type_of_Breach'],axis = 1, inplace = True)
cv = CountVectorizer()
print(x)
print(y)
labeled = 'labeled_data.csv'
data.to_csv(labeled, index=False)
data.to_markdown
x = cv.fit_transform(x)
models = []
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.20)
X_train.shape, X_test.shape, y_train.shape
print("Naive Bayes")
from sklearn.naive_bayes import MultinomialNB
NB = MultinomialNB()
NB.fit(X_train, y_train)
predict_nb = NB.predict(X_test)
naivebayes = accuracy_score(y_test, predict_nb) * 100
print(naivebayes)
print(confusion_matrix(y_test, predict_nb))
print(classification_report(y_test, predict_nb))
```

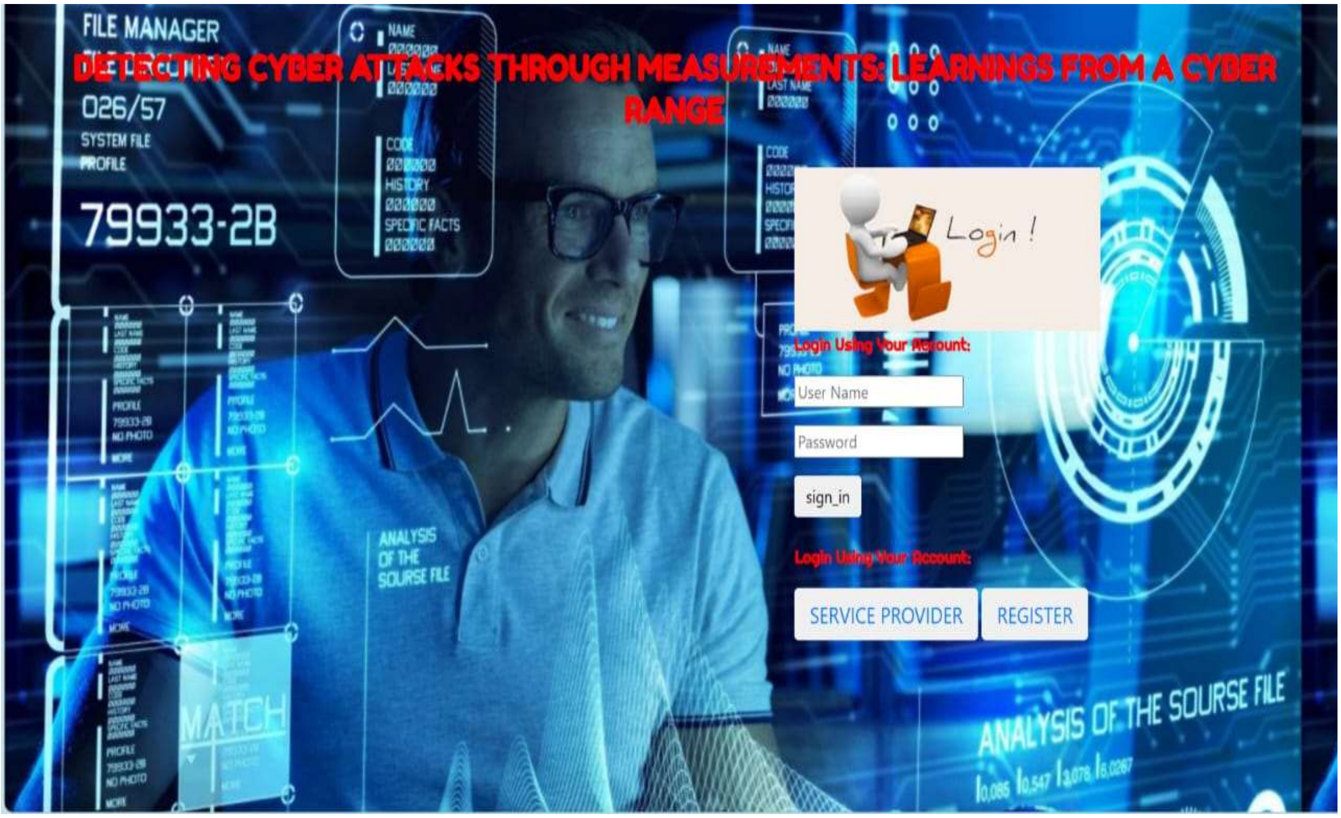
```
models.append(('naive_bayes', NB))
detection_accuracy.objects.create(names="Naive Bayes", ratio=naivebayes)
# SVM Model
print("SVM")
from sklearn import svm
lin_clf = svm.LinearSVC()
lin_clf.fit(X_train, y_train)
predict_svm = lin_clf.predict(X_test)
svm_acc = accuracy_score(y_test, predict_svm) * 100
print(svm_acc)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, predict_svm))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, predict_svm))
models.append(('svm', lin_clf))
detection_accuracy.objects.create(names="SVM", ratio=svm_acc)
print("Logistic Regression")
from sklearn.linear_model import LogisticRegression
reg = LogisticRegression(random_state=0, solver='lbfgs').fit(X_train, y_train)
y_pred = reg.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, y_pred) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, y_pred))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, y_pred))
models.append(('logistic', reg))
detection_accuracy.objects.create(names="Logistic Regression", ratio=accuracy_score(y_test, y_pred) * 100)
print("Decision Tree Classifier")
dtc = DecisionTreeClassifier()
dtc.fit(X_train, y_train)
dtcpredict = dtc.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, dtcpredict) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, dtcpredict))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, dtcpredict))
models.append(('DecisionTreeClassifier', dtc))
 detection_accuracy.objects.create(names="DecisionTreeClassifier",ratio=accuracy_score(y_test,
 dtcpredict) * 100)
print("SGD Classifier")
from sklearn.linear_model import SGDClassifier
sgd_clf = SGDClassifier(loss='hinge', penalty='l2', random_state=0)
sgd_clf.fit(X_train, y_train)
sgdpredict = sgd_clf.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, sgdpredict) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, sgdpredict))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, sgdpredict))
models.append(('SGDClassifier', sgd_clf))
detection_accuracy.objects.create(names="SGD Classifier", ratio=accuracy_score(y_test, sgdpredict) * 100)
print("KNeighborsClassifier")
from sklearn.neighbors import KNeighborsClassifier
kn = KNeighborsClassifier()
kn.fit(X_train, y_train)
knpredict = kn.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, knpredict) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, knpredict))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, knpredict))
models.append(('KNeighborsClassifier', kn))
detection_accuracy.objects.create(names="KNeighborsClassifier", ratio=accuracy_score(y_test, knpredict)
* 100)
```

```
        print("Random Forest Classifier")
        rf = RandomForestClassifier()
rf.fit(X_train, y_train)
pred_rfc = rf.predict(X_test)
print("ACCURACY")
print(accuracy_score(y_test, pred_rfc) * 100)
print("CLASSIFICATION REPORT")
print(classification_report(y_test, pred_rfc))
print("CONFUSION MATRIX")
print(confusion_matrix(y_test, pred_rfc))
      detection_accuracy.objects.create(names="Random Forest Classifier", ratio=accuracy_score(y_test,
     pred_rfc) * 100)
        labeled = 'labeled_data.csv'
        data.to_csv(labeled, index=False)
        data.to_markdown
        obj = detection_accuracy.objects.all()
        return render(request,'SProvider/train_model.html', {'objs': obj})
```

7. SCREENSHOTS

7. SCREENSHOT

7.1 HOME SCREEN



DETECTING CYBER ATTACKS THROUGH MEASUREMENTS: LEARNINGS FROM A CYBER RANGE

Login Using Your Account:

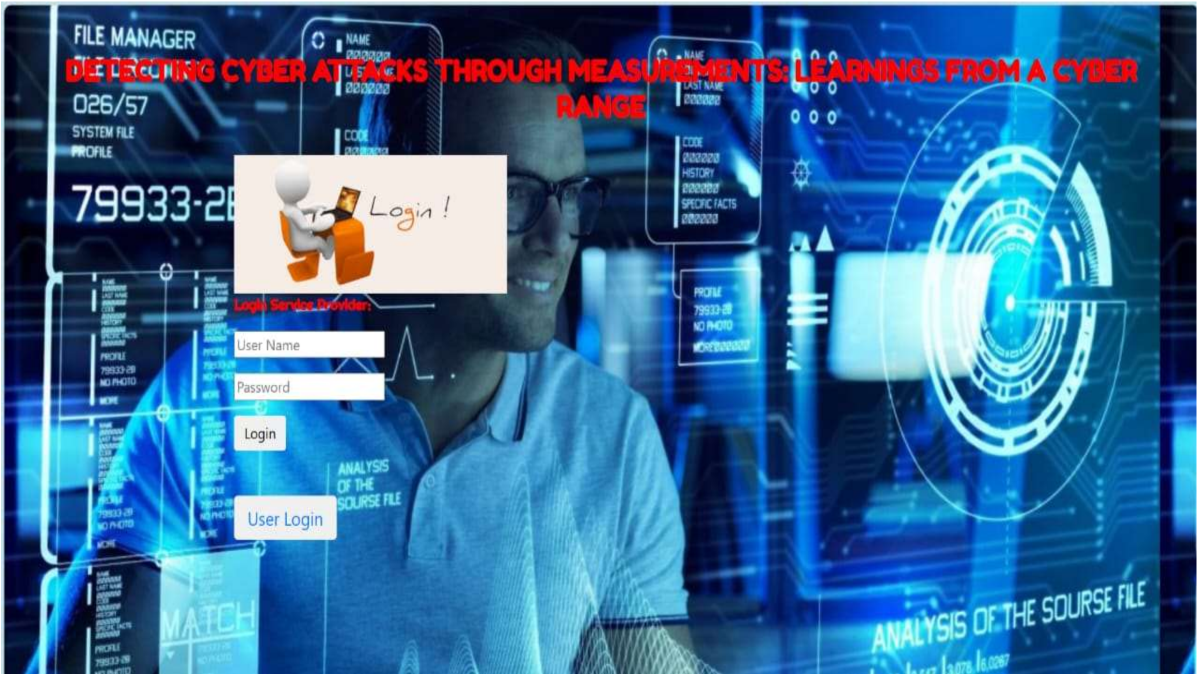
User Name

Password

Login Using Your Account:

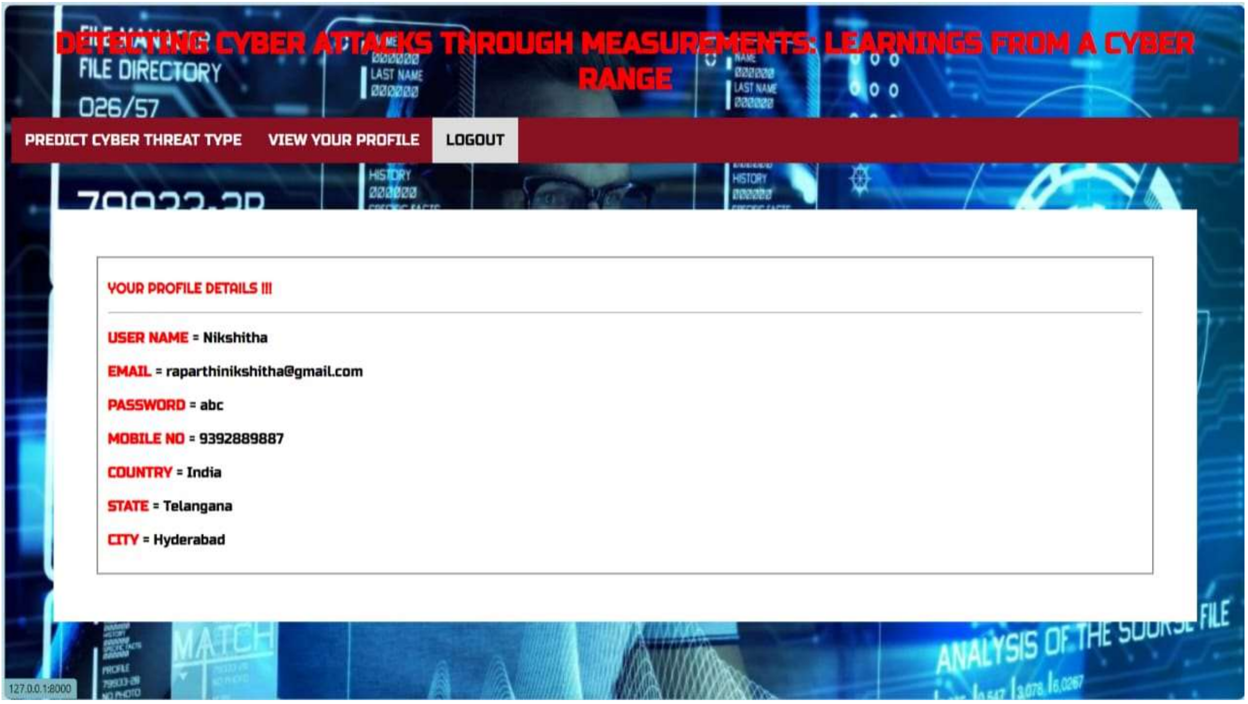
Screenshot 7.1 : HOME SCREEN

7.2 ADMIN PAGE


DETECTING CYBER ATTACKS THROUGH MEASUREMENTS: LEARNINGS FROM A CYBER RANGE
Login !
Login Service Provider:
User Name
Password
Login
User Login

Screenshot 7.2: ADMIN PAGE

7.3 VIEW PROFILE


DETECTING CYBER ATTACKS THROUGH MEASUREMENTS: LEARNINGS FROM A CYBER RANGE
PREDICT CYBER THREAT TYPE
VIEW YOUR PROFILE
LOGOUT
127.0.0.1:8000

YOUR PROFILE DETAILS !!!

USER NAME = Nikshitha

EMAIL = raparthinikshitha@gmail.com

PASSWORD = abc

MOBILE NO = 9392889887

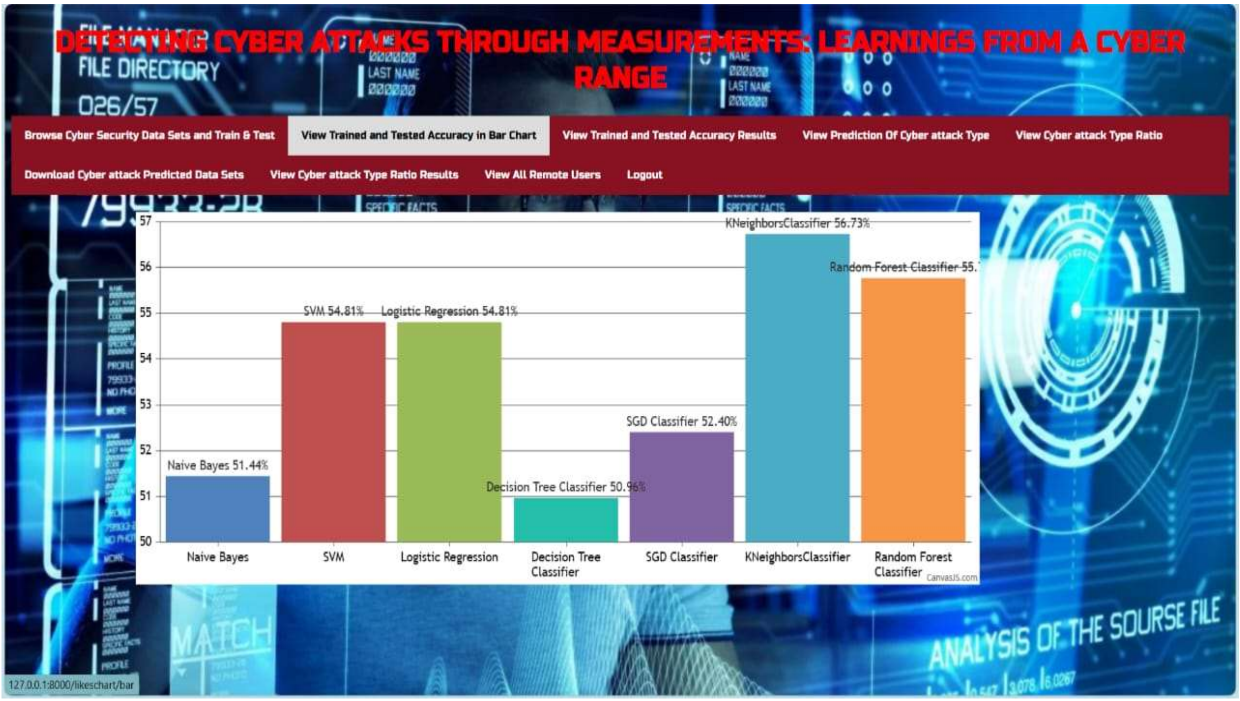
COUNTRY = India

STATE = Telangana

CITY = Hyderabad

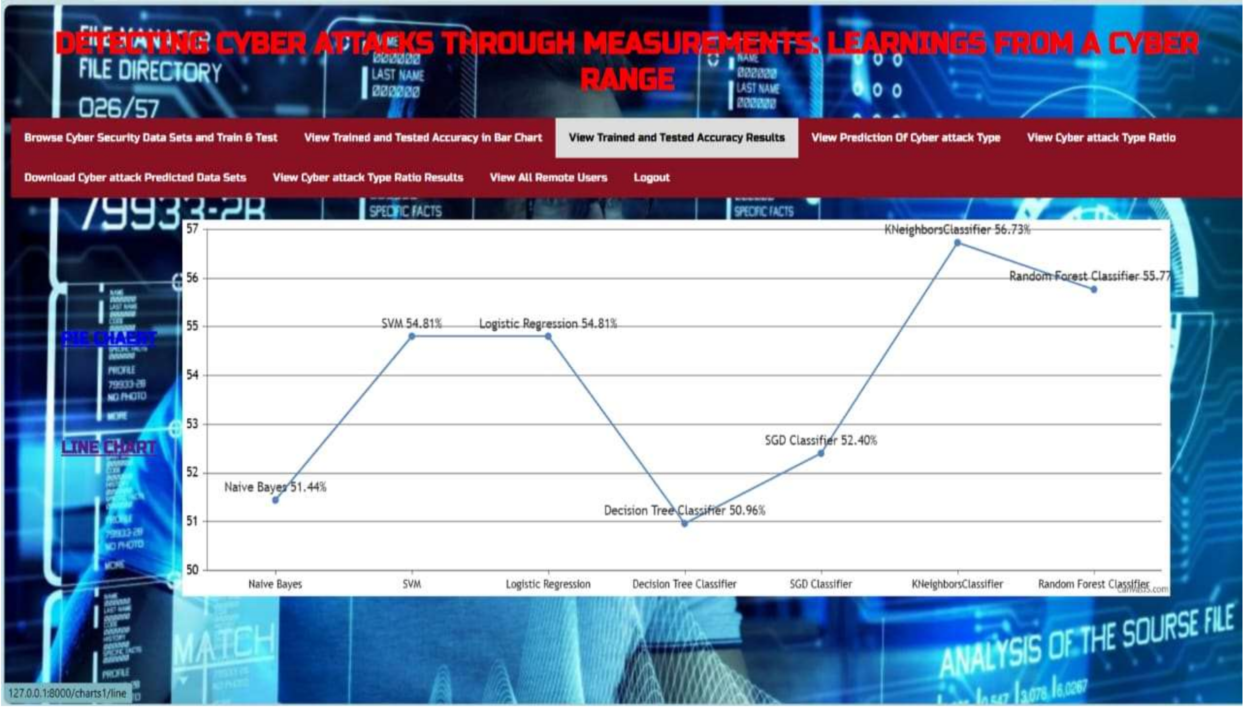
Screenshot 7.3 : VIEW PROFILE

7.4 TRAINED AND TEST RESULT


DETECTING CYBER ATTACKS THROUGH MEASUREMENTS: LEARNINGS FROM A CYBER RANGE
Browse Cyber Security Data Sets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Cyber attack Type
View Cyber attack Type Ratio
Download Cyber attack Predicted Data Sets
View Cyber attack Type Ratio Results
View All Remote Users
Logout
57
56
55
54
53
52
51
50
Naive Bayes 51.44%
SVM 54.81%
Logistic Regression 54.81%
Decision Tree Classifier 50.96%
SGD Classifier 52.40%
KNeighborsClassifier 56.73%
Random Forest Classifier 55.7
Naive Bayes
SVM
Logistic Regression
Decision Tree Classifier
SGD Classifier
KNeighborsClassifier
Random Forest Classifier
CanvasJS.com
127.0.0.1:8000/likeschart/bar

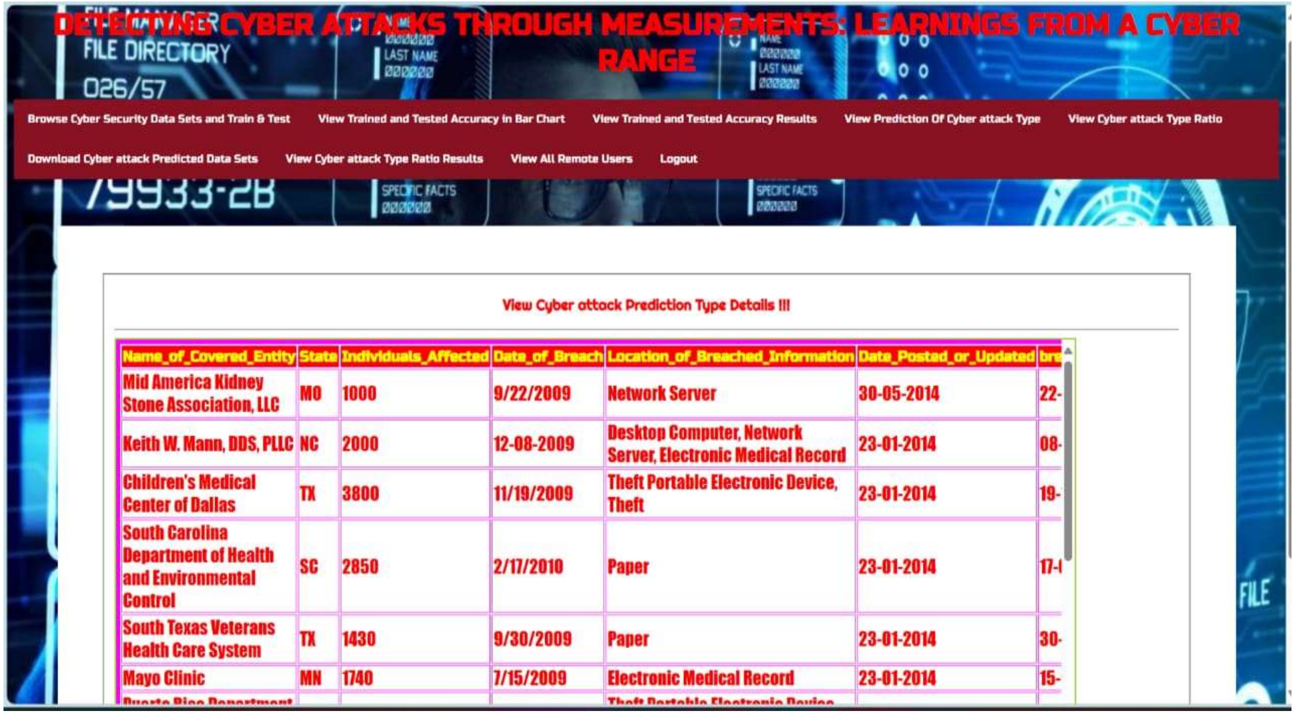
Screenshot 7.4 :TRAINED AND TEST RESULTS

7.5 TRAINED AND TESTED ACCURACY RESULTS


DETECTING CYBER ATTACKS THROUGH MEASUREMENTS: LEARNINGS FROM A CYBER RANGE
Browse Cyber Security Data Sets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Cyber attack Type
View Cyber attack Type Ratio
Download Cyber attack Predicted Data Sets
View Cyber attack Type Ratio Results
View All Remote Users
Logout
PIE CHART
LINE CHART
57
56
55
54
53
52
51
50
Naive Bayes 51.44%
SVM 54.81%
Logistic Regression 54.81%
Decision Tree Classifier 50.96%
SGD Classifier 52.40%
KNeighborsClassifier 56.73%
Random Forest Classifier 55.77
Naive Bayes
SVM
Logistic Regression
Decision Tree Classifier
SGD Classifier
KNeighborsClassifier
Random Forest Classifier
CanvasJS.com
127.0.0.1:8000/charts1/line

Screenshot 7.5: TRAINED AND TEST ACCURACY RESULTS

7.6 PREDICTION OF TYPES OF CYBER ATTACKS

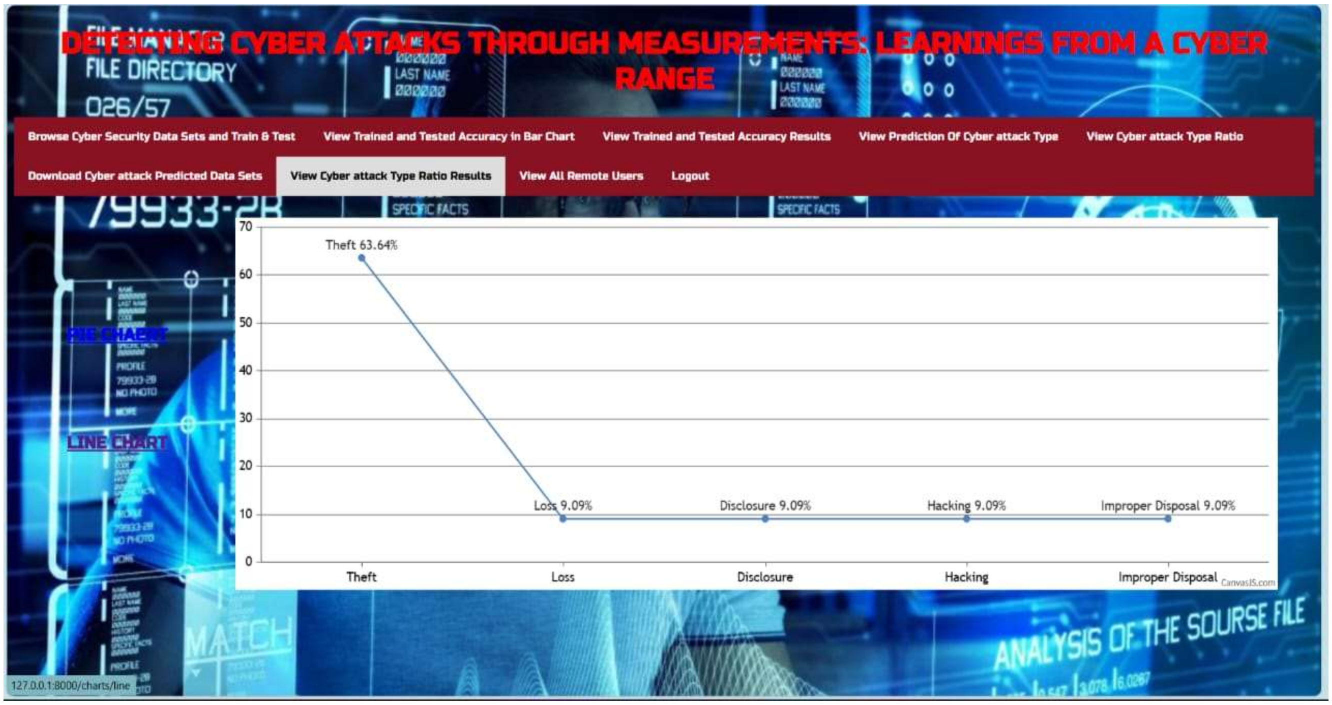

DETECTING CYBER ATTACKS THROUGH MEASUREMENTS: LEARNINGS FROM A CYBER RANGE
Browse Cyber Security Data Sets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Cyber attack Type
View Cyber attack Type Ratio
Download Cyber attack Predicted Data Sets
View Cyber attack Type Ratio Results
View All Remote Users
Logout

View Cyber attack Prediction Type Details !!!

Name_of_Covered_Entity	State	Individuals_Affected	Date_of_Breach	Location_of_Breached_Information	Date_Posted_or_Updated	bre
Mid America Kidney Stone Association, LLC	MO	1000	9/22/2009	Network Server	30-05-2014	22-
Keith W. Mann, DDS, PLLC	NC	2000	12-08-2009	Desktop Computer, Network Server, Electronic Medical Record	23-01-2014	08-
Children's Medical Center of Dallas	TX	3800	11/19/2009	Theft Portable Electronic Device, Theft	23-01-2014	19-
South Carolina Department of Health and Environmental Control	SC	2850	2/17/2010	Paper	23-01-2014	17-(
South Texas Veterans Health Care System	TX	1430	9/30/2009	Paper	23-01-2014	30-
Mayo Clinic	MN	1740	7/15/2009	Electronic Medical Record	23-01-2014	15-
Puerto Rico Department				Theft Portable Electronic Device		

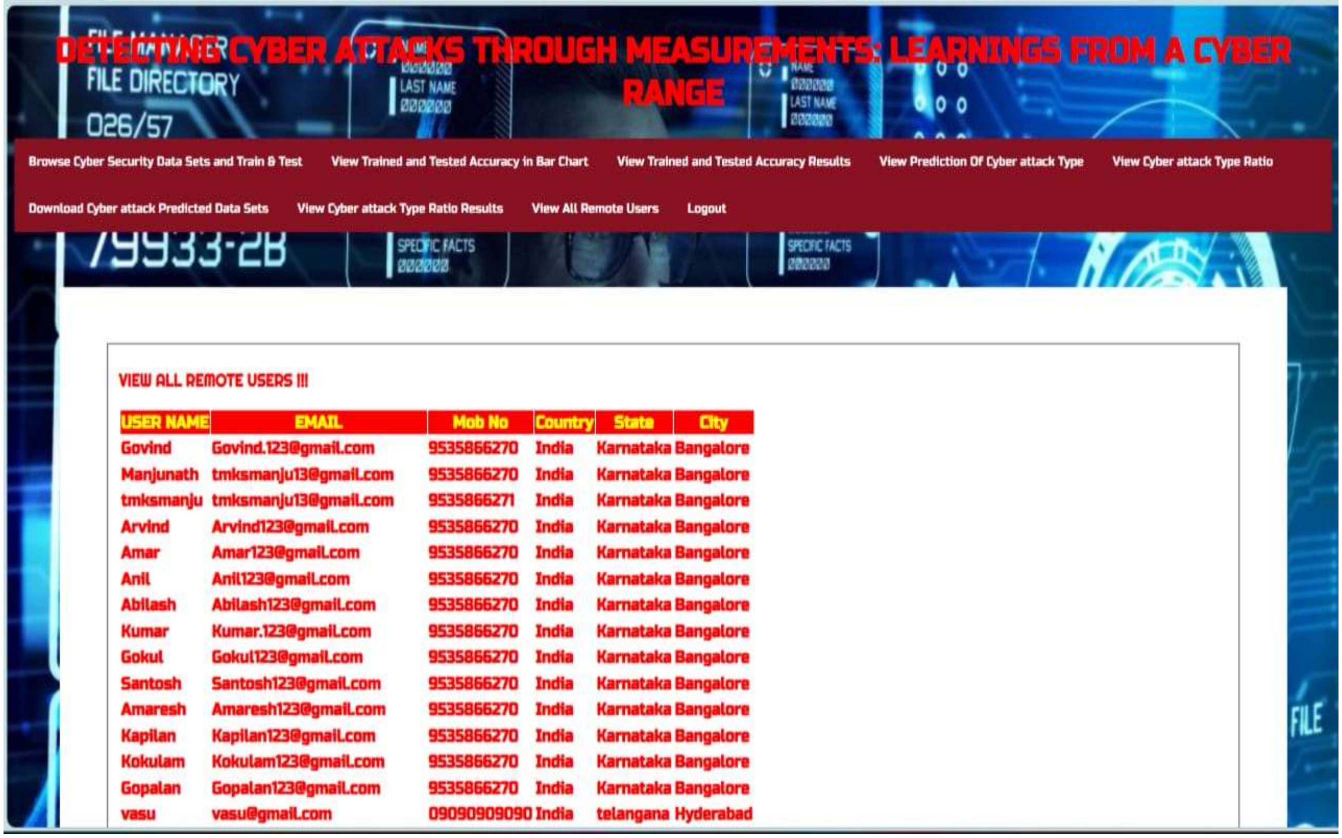
Screenshot 7.6: PREDICTION OF TYPES OF CYBER ATTACKS

7.7 VIEW CYBER ATTACK TYPE RATIO



Screenshot 7.7 :VIEW CYBER ATTACK TYPE RATIO

7.8 VIEW ALL REMOTE USERS


DETECTING CYBER ATTACKS THROUGH MEASUREMENTS: LEARNINGS FROM A CYBER RANGE
Browse Cyber Security Data Sets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Cyber attack Type
View Cyber attack Type Ratio
Download Cyber attack Predicted Data Sets
View Cyber attack Type Ratio Results
View All Remote Users
Logout

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Mob No	Country	State	City
Govind	Govind.123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	9535866270	India	Karnataka	Bangalore
tmksmanju	tmksmanju13@gmail.com	9535866271	India	Karnataka	Bangalore
Arvind	Arvind123@gmail.com	9535866270	India	Karnataka	Bangalore
Amar	Amar123@gmail.com	9535866270	India	Karnataka	Bangalore
Anil	Anil123@gmail.com	9535866270	India	Karnataka	Bangalore
Abilash	Abilash123@gmail.com	9535866270	India	Karnataka	Bangalore
Kumar	Kumar.123@gmail.com	9535866270	India	Karnataka	Bangalore
Gokul	Gokul123@gmail.com	9535866270	India	Karnataka	Bangalore
Santosh	Santosh123@gmail.com	9535866270	India	Karnataka	Bangalore
Amaresh	Amaresh123@gmail.com	9535866270	India	Karnataka	Bangalore
Kapilan	Kapilan123@gmail.com	9535866270	India	Karnataka	Bangalore
Kokulam	Kokulam123@gmail.com	9535866270	India	Karnataka	Bangalore
Gopalan	Gopalan123@gmail.com	9535866270	India	Karnataka	Bangalore
vasu	vasu@gmail.com	09090909090	India	telangana	Hyderabad

Screenshot 7.8 : VIEW ALL REMOTE USERS

8.TESTING

8. TESTING

8.1 PURPOSE OF TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discoverevery conceivable fault or weakness in a work product. It provides a way to check thefunctionality of components, sub-assemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail unacceptably. There are various types of tests. Each test type addresses a specific testing requirements.

8.2 TYPES OF TESTING

8.2.1 UNIT TESTING

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application

It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basictests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately tothe documented specifications and contains clearly defined inputs and expected results.

8.2.2 INTEGRATION TESTING

Integration tests are designed to test integrated software components to determine if theyactually run as one program. Testing is event driven and is more concerned with the basic out come of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components iscorrect and consistent.

Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

8.2.3 FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available asspecified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.
- Systems/Procedures : interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows;data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value ofcurrent tests is determined.

8.2.4 SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points

8.2.5 WHITE BOX TESTING

White Box Testing is a testing in which in which the software tester has knowledge ofthe inner workings, structure and language of the software, or at least its purpose. It is purpose. I is used to test areas that cannot be reached from a black box level

8.2.6 BLACK BOX TESTING

This Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds oftests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

8.2.7 UNIT TESTING

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as twodistinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written

Test objectives

All field entries must work properly.
Pages must be activated from the identified link.
The entry screen, messages and responses must not be delayed.

Features to be tested

Verify that the entries are of the correct format

No duplicate entries should be allowed

All links should take the user to the correct page

INTEGRATION TESTING

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.The taskof the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

ACCEPTANCE TESTING

User Acceptance Testing is a critical phase of any project and requires significant participation by theend user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered

9. CONCLUSION AND FUTURE SCOPE

9. CONCLUSION AND FUTURE SCOPE

9.1 CONCLUSION

This article discusses the objectives of an SOC for monitoring an infrastructure as well as presents the ELK-Stack and its use in this context,while the different Log sources according to their information and priority are presented. In addition, ELK-Stack in action is showcased, where a cyber was able to be detected and the attacker's intentions were mapped. During the game scenario, we are able to log network traffic, authentication attempts, used commands, files and accessed applications. These logs are an indicator of the malicious actions to gain control of a system by guessing its password, and then how it is used as a step-ping stone to attack other devices. In our implementation, the ELK-Stack is shown to be effective both for log collection and organization, and also for detecting a cyber incident.For our future work, we intend to use the collected data, for research purposes and for deriving intelligence and combine it with real attacks indicators. That can contribute towards solving the ground truth problem by having the knowledge of what incidents occurred in the system.From a real-world perspective, organizations can share their attack data, namely IoCs with the community. For instance, considering a malicious activity, the organization can share the IP addresses and domains involved, or the files and binaries footprints left by the attackers during the activity. This approach could spread the awareness on the community and help to reduce the risk for other organizations, assuming that the attackers are using the same tools and techniques.

9.2 FUTURE SCOPE

In the future, the project's scope could be expanded to include more advanced machine learning techniques and data analysis methods to further enhance the detection system's accuracy and efficiency. Additionally, incorporating threat intelligence feeds and integrating the system with other security tools could improve its capabilities in detecting and responding to emerging cyber threats. Furthermore, expanding the training program to include more comprehensive cybersecurity concepts and hands-on exercises could better prepare professionals to handle complex cyber attacks in real-world scenarios.

10.BIBLIOGRAPHY

10.BIBILOGRAPHY

10.1 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, “Security operations center: a systematic study and open challenges,” IEEE Access, 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] G. González-Granadillo, S. González-Zarzosa, and R. Diaz, “Security information and event management (SIEM): analysis, trends, and usage in critical infrastructures,” Sensors, vol. 21, no. 14, p. 4759, Jul. 2021.

[9] R.-V. Mahmoud, E. Kidmose, A. Turkmen, O. Pilawka, and J. M. Pedersen, ‘DefAtt architecture of virtual cyber labs for research and education,” in Proc. 2021 Int. Conf. Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), pp. 1–7, Jun. 2021