Department of Information Technology

A.P. Shah Institute of Technology

G.B.Road,Kasarvadavli, Thane(W),Mumbai-400615

UNIVERSITY OF MUMBAI

Academic Year 2020-2021

A Project Report on

AI Based SOAR

Submitted in partial fulfillment of the degree of
Bachelor of Engineering(Sem-8)
in

INFORMATION TECHNOLOGY

By
Rahul Vast (17104042)
Shruti Sawant(18204001)
Aishwarya Thorbole(18204002)

Under the Guidance of
Mr. Vishal Badgujar

1.Project Conception and Initiation

Cybersecurity is becoming very crucial in the today's world where technology is now not limited to just computers, smartphones, etc. It is slowly entering into things that are used on daily basis like home appliances, automobiles, etc. Thus, opening a new door for people with wrong intent. With the increase in speed of technology dealing with such issues also requires quick response from security people. In environments that deal with a high volume of events, analysts often spend a significant amount of time resolving these security alerts.

1.1 Abstract

- In today's world where technology is now not limited to just computers, smart phones, etc.
- It is slowly entering into things that are used on a daily basis like home appliances, automobiles, etc. Thus, opening a new door for people with wrong intent.
- With the increase in speed of technology dealing with such issues also requires quick response from security people. Thus, dealing with a huge variety of devices quickly will require some extent of automation in this field.
- For this the very first step is that the collected data from different sources will be converted into a standardized format i.e. to categorize the data collected from different sources.

1.2 Objectives

- To translate a threat intelligence available in different languages.
- To collect data from different data sources and represent them in a standardized format so they can be easily classified.
- To recognize different patterns from the event logs/data-set, detection of the vulnerabilities the loop holes in the system and prevent them.
- To perform event profiling such that it will help to find out the attack patterns as well as to identify the pattern while monitoring.
- To differentiate between false positive and true positive types of alerts-identification of the actual alert and elimination of the false positive one.
- To identify almost all form of attacks and respond to attacks like Malware Attacks, SQL Injection,MITM,Password Attacks and DDOS.

1.3 Literature Review

Sr No	Author Names	Methodology	Advantages	Disadvantages	IEEE Conference	Year Of Publication	Our Findings
1	Jonghoon Lee,Jonghyun Kim,Ikkyun Kim and Kijun Han	Author suggested an AI technique for Cyber-Threats Detection based on artificial neural networks.Also an event profiling method for pre-processing is suggested	The Complexity can be reduced and bias toward frequent records by machine learning algorithms can be prevented	The model suggested by author requires high computational power and storage as well. Consistent tuning and daily updates according to modern threats is requiring human assistance.	IEEE Access	2019	The AI-SIEM system Enables the security analysts to deal with significant security alerts promply .

Sr No	Author Names	Methodology	Advantages	Disadvantages	IEEE Conference	Year Of Publication	Our Findings
2	Hamad AL-Mohannadi,Irfan Awan ,Jassim Al Hamar,Andrea Cullen,Jules Pagan Disso and Lorna Armitage	Author suggested anew threat Intelligence technique which is evaluated by analyzing honeypot log data to identify behavior of attackers to find attack patterns. Honeypot data analysis is the way by which one can detect and understand cyber threats.	Using honeypots data for threat intelligence is that there is no side effect on the production system.	The potential drawback of this Model is that it requires data related to same incident in huge amount.	IEEE 32nd International Conference On advanced information networking and appliances(AINA)	2018	These system Generates Alerts and prevent Cyber Attacks
3	Farhan Sadique,Sui Cheung,Iman Vakilinia,Shahriar Badsha,Shamik Sengupta	Author suggested a mechanism to represent raw cyber threat-data in Structured Threat Information Expression format in an automated manner and also took care of privacy preservation.	They Improve the Privacy of The Data Because Sensitive information is removed as a CTI is generated.	-	IEEE Annual Ubiquitous Computing Electronics and mobile communication Conference (UEMCON)	2018	The Standard Format required for an organization to share these data with other Organizations is provided with these systems.

Sr No	Author Names	Methodology	Advantages	Disadvantages	IEEE Conference	Year Of Publication	Our Findings
4	Priyanka Ranade,Sudip Mittal,Anupam Joshi And Karuna Joshi	Author suggested to create a neutral network that takes in threat intelligence available in different languages using which it translates that in desired languages and thus help in making available threat intelligence in different language.	The system is able to run independently in secluded operational settings.	The System requirement of a cybersecurity rich alignment to train the model is quite high though.	IEEE International Conference on Intelligence and security Informatics(ISI)	2018	This proposed system uses Russian And English word embeddings created from cybersecurity data.

1.4 Problem Definition

• Traditional Methods for Threat Detection have the inability to gather appropriate/relevant data to be analyzed for true positives. Also these methods lack quick response mechanisms for preventing modern attacks.

• IDS may not be potentially always correct it may include much more false positives. Here a new threat intelligence technique evaluated by analyzing honeypot data to identify attackers behaviour to find attack patterns.

• The multilingual nature of the Internet increases complications in the cybersecurity community's ongoing efforts to strategically mine threat intelligence from sources such as social media blogs and dark web vulnerability markets that exists in the diverse languages that security analysts often hinder about.

1.5 Scope

•A SIEM (Security Information and Event Management) system lowers the burden of security teams by managing most of the daily repetitive things and organizing work.

•This technology is getting more attention from cybersecurity field as this technology has more potential in future.

•It is also getting evolved in the form of SOAR.

•Threat intelligence holds great importance in the field of cyber security.

•It helps the incident response teams to have intel on the threat occurring as well as the well-known tested methods to prevent it. Thus, the below method has proposed a way to automate the threat intelligence gathering process along with privacy preservation keeping in mind.

1.6 Technology Stack

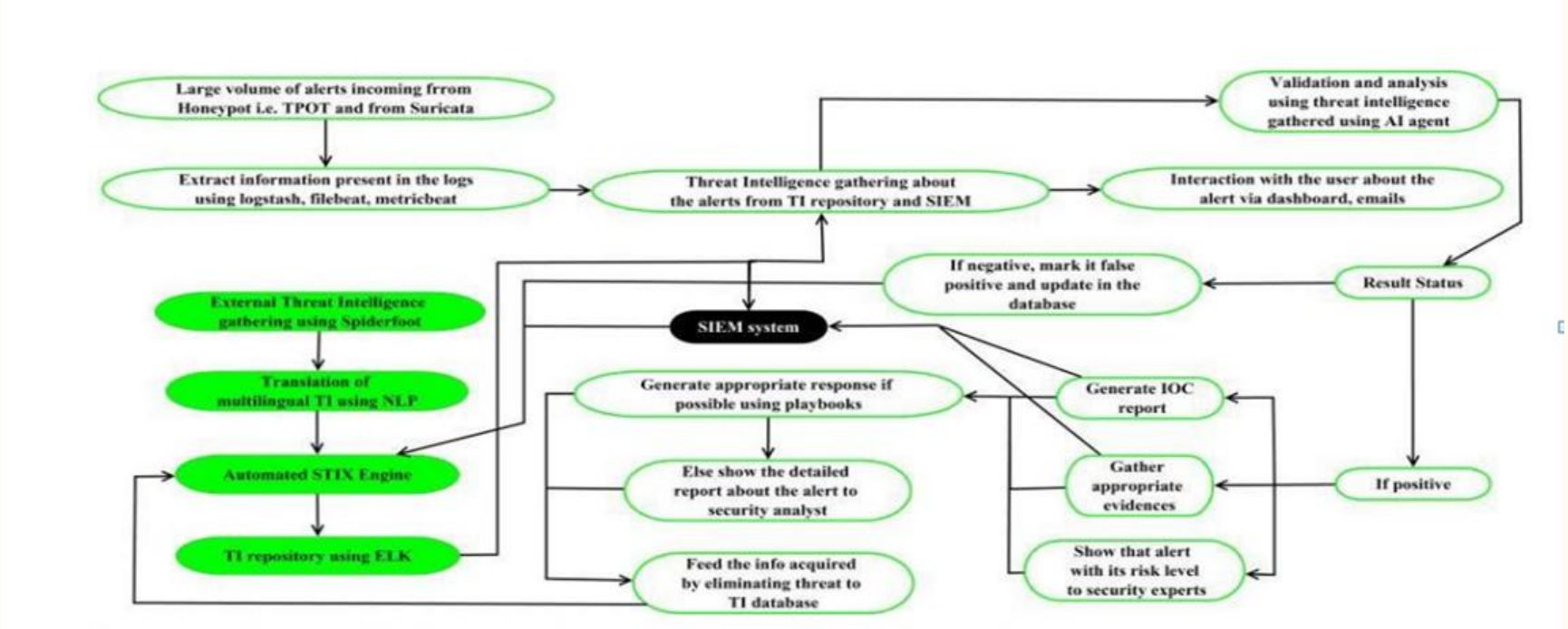
- Suricata for Network Security Python for programming the business logic
- AWS for computational power, storage and deployment
- ELK Stack for data visualizations and analysis
- Red hat Ansible for automating configurations
- Docker for creating container and ease deployment
- Spider foot /Virus Total API for automated of Monitoring Engine
- Cockpit for server administration
- Tpotce for honeypot
- Light GBM(Gradient Booster Model)
- STIX-2
- Ember

1.7 Benefits for environment & Society

- RAPID RESPONSE- Going beyond data acquisition and analysis, SOAR solutions can be configured to respond automatically to a range of situations
- CONSISTENCY AND COMPLIANCE- Automation eliminates the possibility of human error and reduces the number of judgement calls analysts are required to make.
- FOCUSED ATTENTION- By automating the handling of these alerts, analysts can devote more of their time and attention to situations where human intervention really is required while the software handles the rest.

2. Project Design

2.1 Proposed System



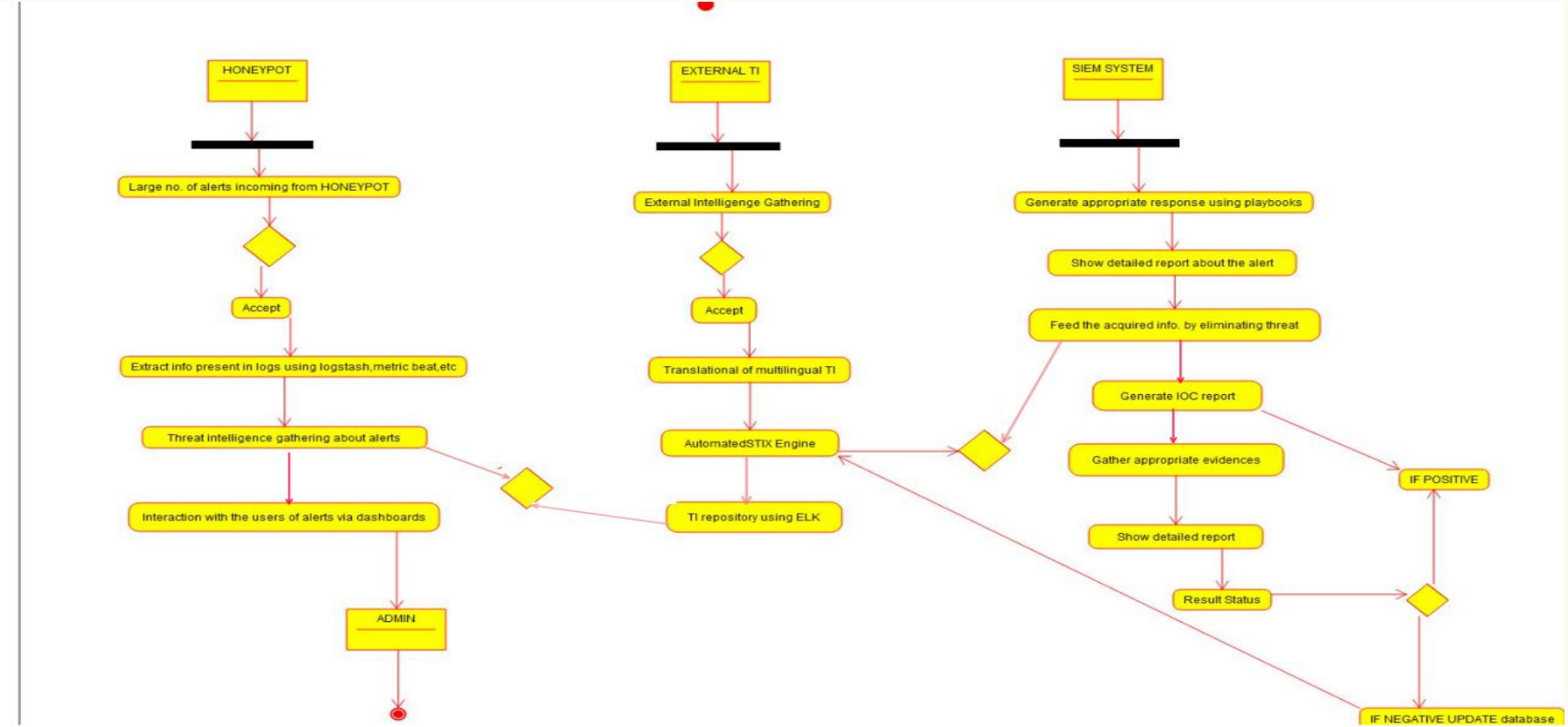
2.1 AI Based Security Orchestration, Automation and Response

It focus of this technology is to Automate various security processes like network security audit, privileged password management and coordination and execution of tools between various tools and security groups.

2.2 Description Of Use Case

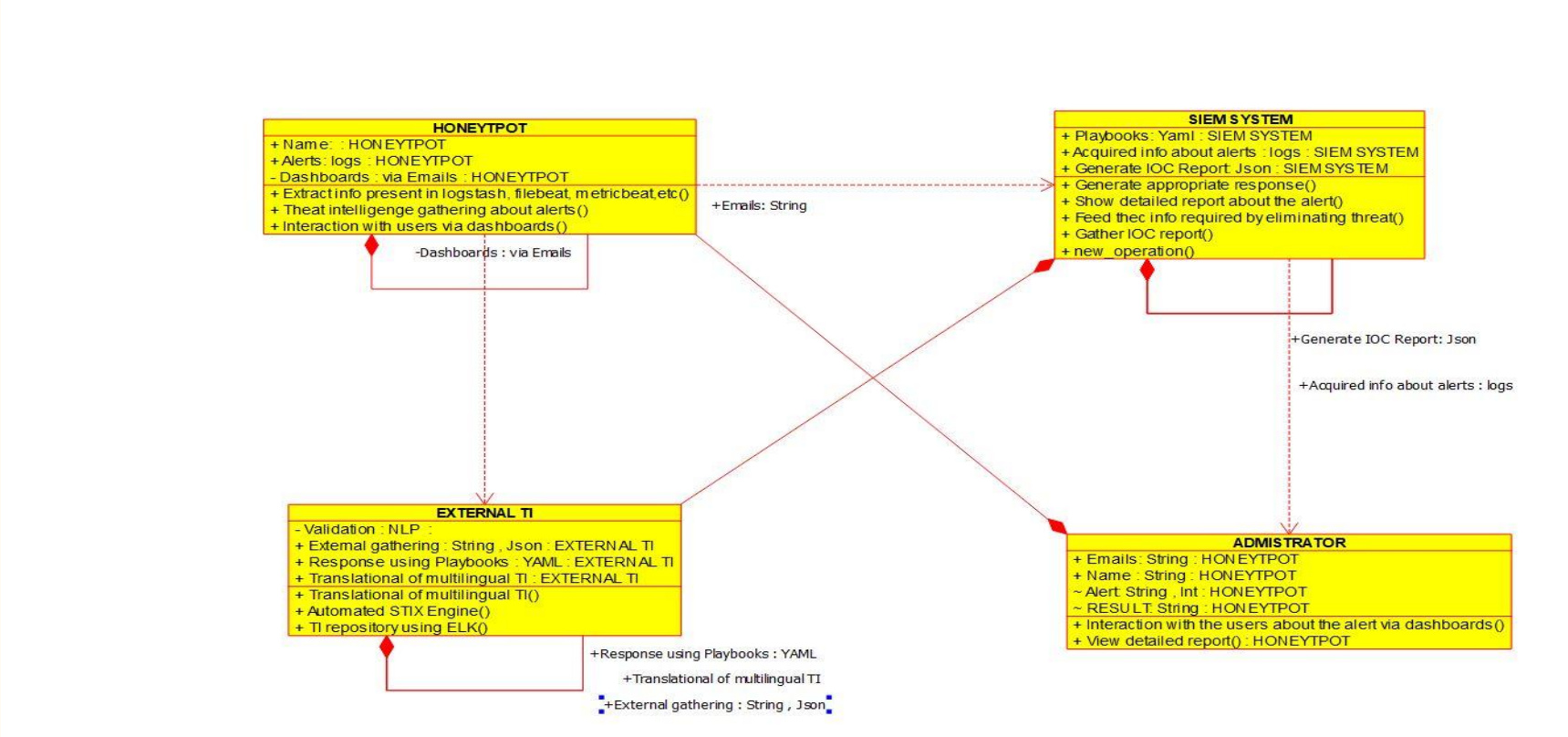
2.2 Use case Diagram for AI BASED SOAR

2.3 Activity diagram



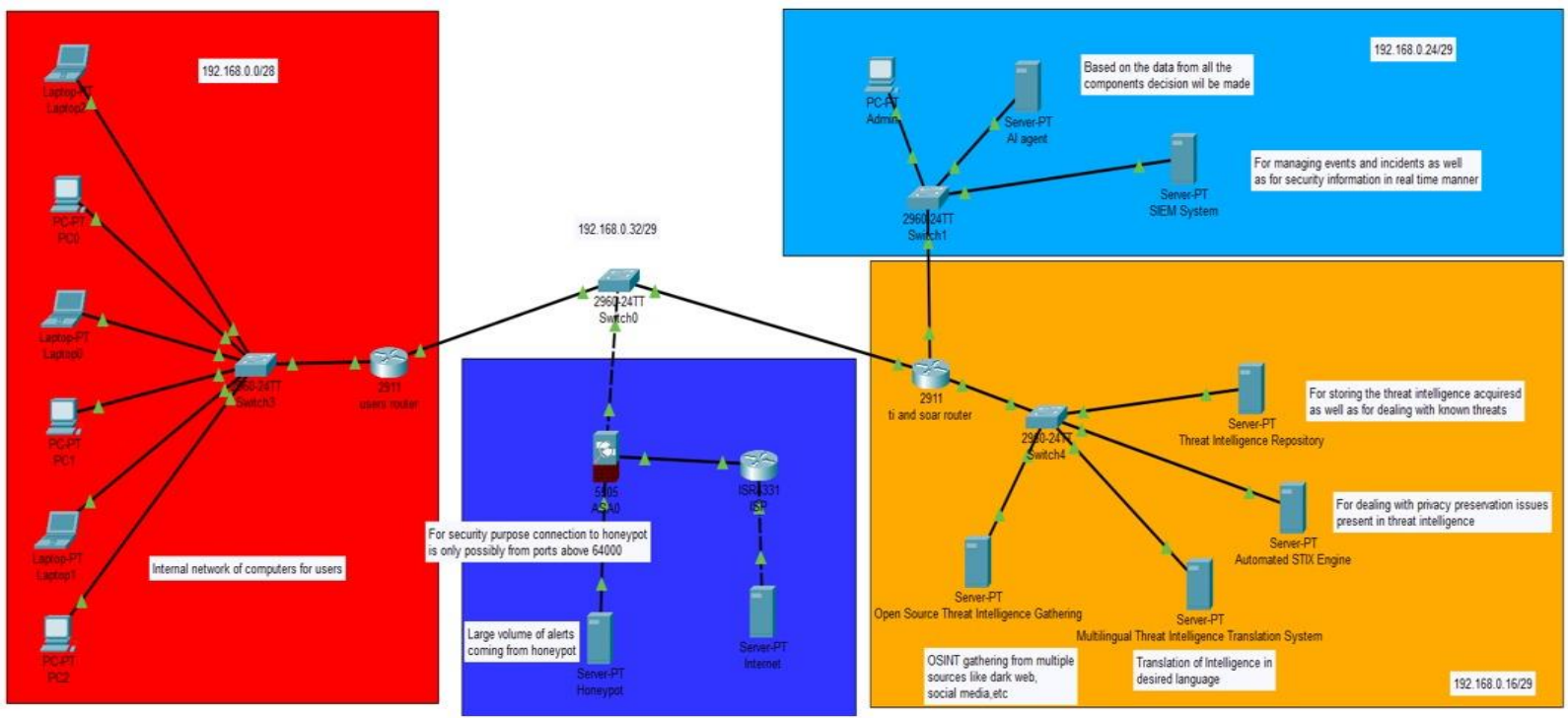
2.3 Activity Diagram for AI BASED SOAR

2.4 Class Diagram



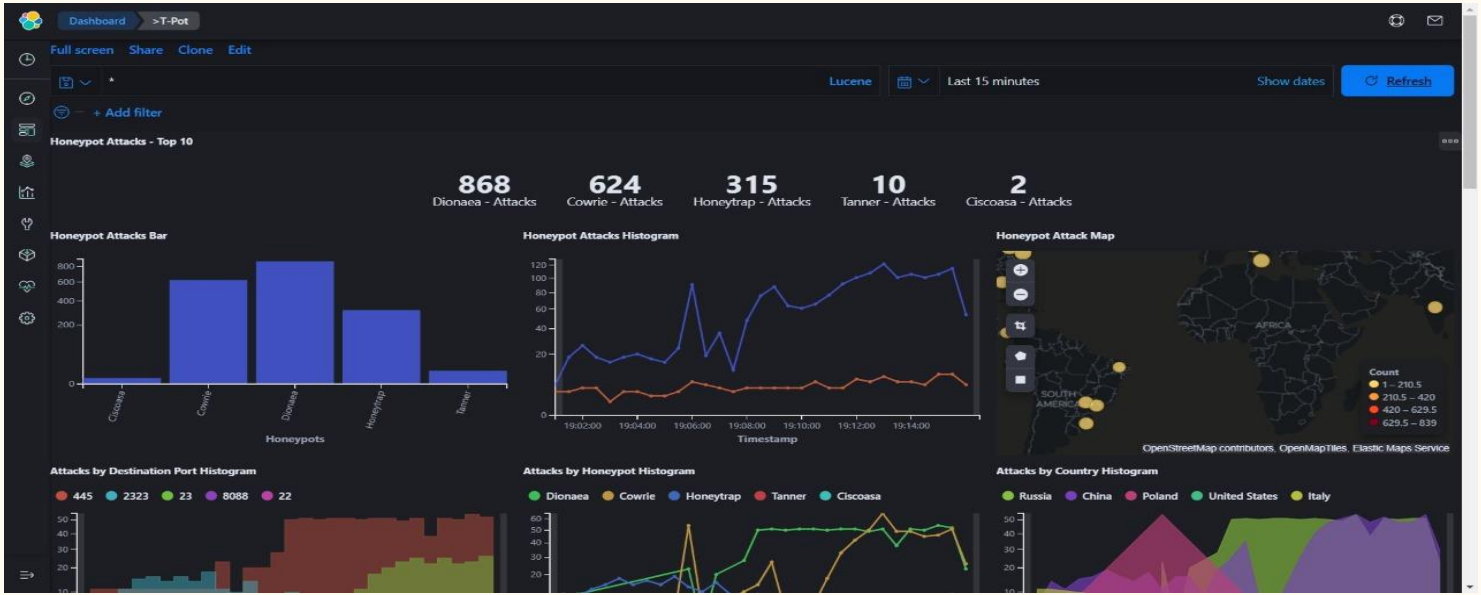
2.4 Class Diagram for AI BASED SOAR

2.5 Module-1



2.5 Cisco Packer Tracker Topology

2.5.1 Module 2



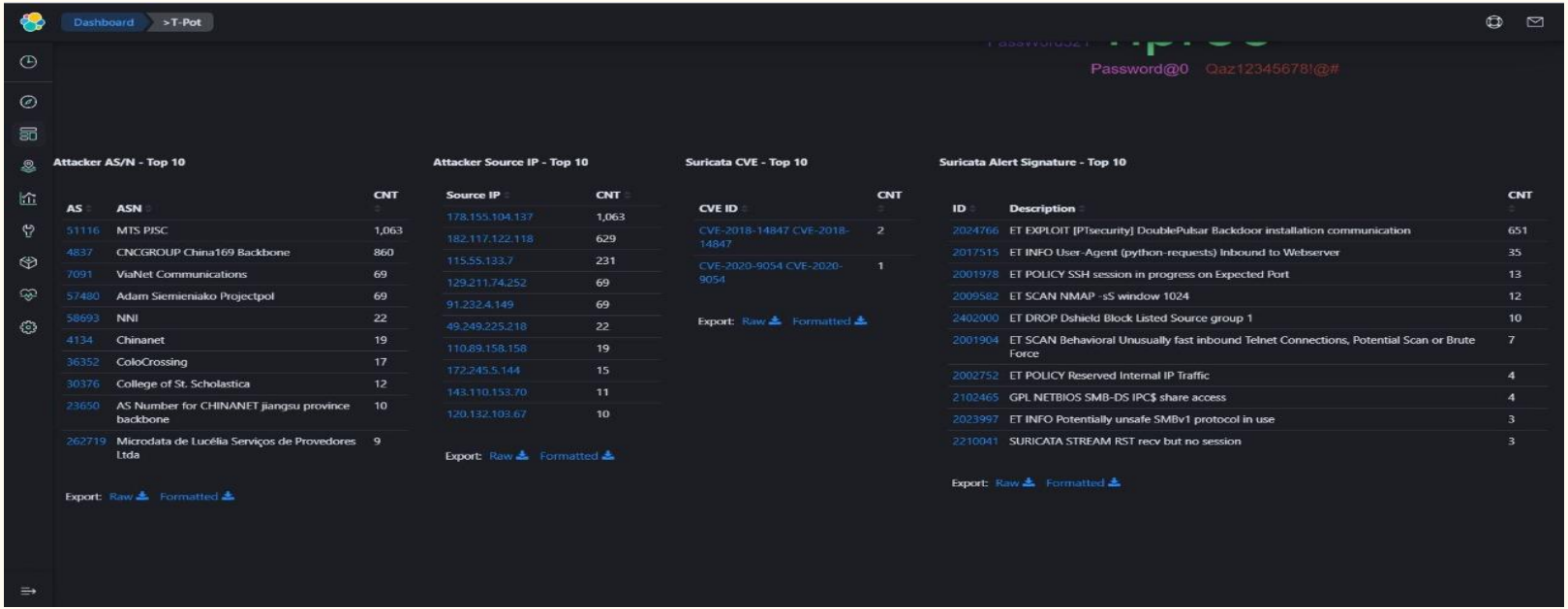
2.5.1 Honeyypot Attacks

The Above is the summarized in honey pot attack Bar graph, honeypot attacks by histogram, attacks by country histogram and so on.

2.5.2 Module 3

2.5.2 Attacks by Destination Port Histogram Being depicted by using histogram in which has Attacker SRC(source) IP Reputation, attacks by honeypot where attacks Dionaea, Cowrie, Tanner etc.

2.5.3 Module 4



2.5.3 Attacker Dashboard T-Pot

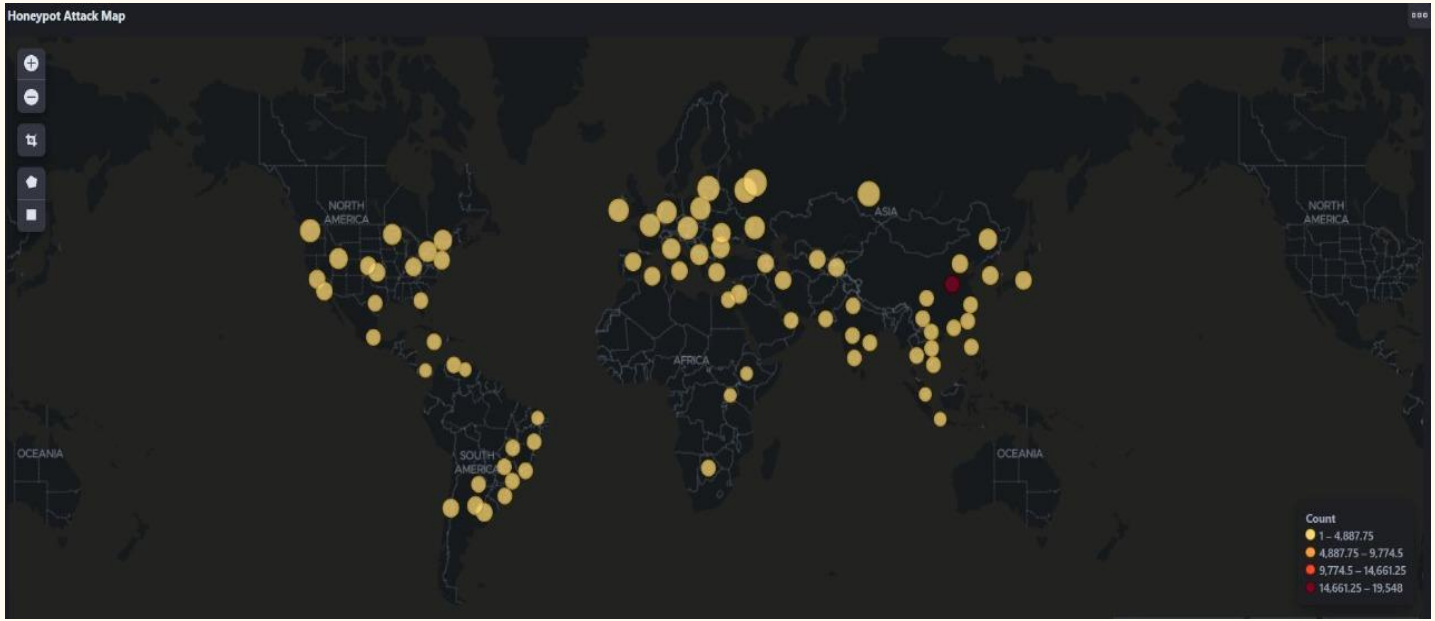
The above figure, depicts the different attacks there source IPs and its count.

2.5.4 Module 5

2.5.4 Honeypot Attacks

Above is the summarized the honeypot attacks by histogram, attacks by country histogram and so summarized in honeypot attack Bargraph honeypot.

2.5.5 Module 6



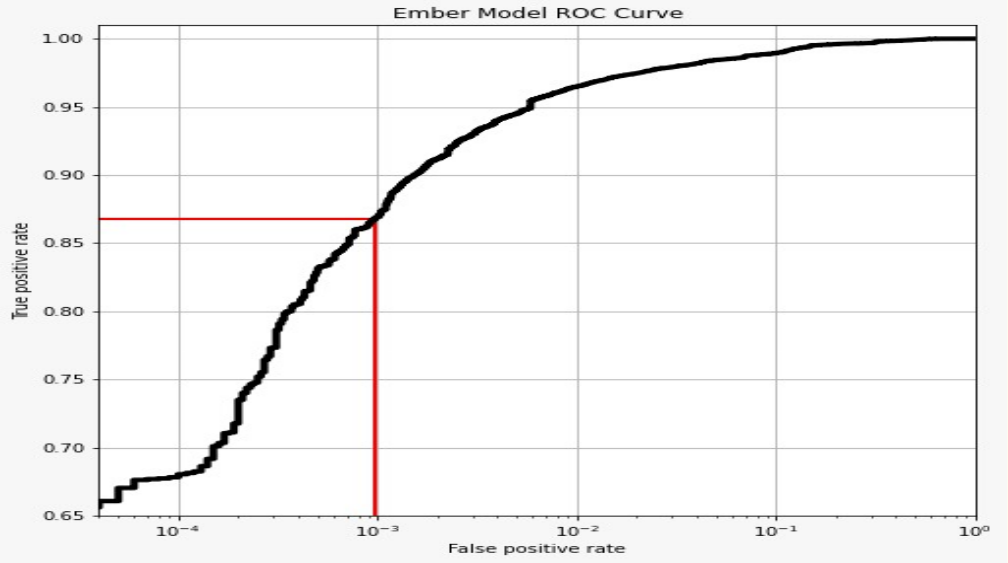
2.5.5 Honeypot Attack Map

Above is the Honeypot Attack Map with the satellite view of countries with most attack, the red dot has the maximum attacks as seen.

2.5.6 Module 7

2.5.6 Attacks by Country and Port
Here the attacks are categorized by country and port where the different country's attack count can be seen.

2.5.7 Module 8



2.5.7 The Above Curve plots true positive and false positive attack rate

2.5.8 Module 9

2.5.8 Wannacry
Threat intelligence of Wannacry run model

2.5.9 Module 10

```
{'Modified': '2018-10-12T21:58:00',
 'Published': '2010-11-10T03:00:00',
 'access': {'authentication': 'NONE',
            'complexity': 'MEDIUM',
            'vector': 'NETWORK'},
 'assigner': 'cve@mitre.org',
 'capec': [{'id': '46',
            'name': 'Overflow Variables and Tags',
            'prerequisites': 'The target program consumes user-controllable '
                             'data in the form of tags or variables. The '
                             'target program does not perform sufficient '
                             'boundary checking.',
            'related_weakness': ['118',
                                 '119',
                                 '120',
                                 '20',
                                 '680',
                                 '697',
                                 '733',
                                 '74'],
            'solutions': 'Use a language or compiler that performs automatic '
                         'bounds checking. Use an abstraction library to '
                         'abstract away risky APIs. Not a complete solution. '
                         'Compiler-based canary mechanisms such as StackGuard, '
                         'ProPolice and the Microsoft Visual Studio /GS flag. '
                         'Unless this provides automatic bounds checking, it '
                         'is not a complete solution. Use OS-level '
                         'preventative functionality. Not a complete solution. '
                         'Do not trust input data from user. Validate all user '
                         'input.',
            'summary': 'This type of attack leverages the use of tags or '
                       'variables from a formatted configuration data to cause '
                       'buffer overflow. The attacker crafts a malicious HTML '
                       'page or configuration file that includes oversized '
                       'strings, thus causing an overflow.'},
```

2.5.9 Vulnerability Response

It Showcases this the TI which contains the entities such as name,assigners,prereqisites,related-weaknesses,solutions,summary and so on.

3.Conclusion and Future Scope

- In our work, we are proposing a system that will bring a certain amount of automation with the help of emerging technologies like A.I that will help to reduce the burden of day to day activities of security professionals like going through the millions of logs swiftly, carrying out necessary procedures and targeting the areas where human intervention is required the most. We are also planning to automate one of the important factors that plays major role in identifying and preventing previously held attacks i.e. Threat Intelligence collection. Also, we are trying to translate threat intelligence in different languages. In future work, we will be focusing more on the zero-day exploit attacks, drive-by attacks and the eavesdropping attacks. Preventing these will be a need in the future as such attacks are getting increased day by day. A.I. based Security, Orchestration, Automation and Response System Workflow (Proposed System) 4 translation engine for multilingual threat intelligence will also be made for other languages such as Japanese and French languages. This will help us to increase the reach of our system.

4.References

[1] Farhan Sadique, Raghav Kaul, Shahriar Badsha, Shamik Sengupta, “An Automated Framework for Real-time Phishing URL Detection”, in IEEE 10th Annual Computing and Communication Workshop and Conference (CCWC), Accepted For Publications, 2020.

[2] Farhan Sadique, Khalid Bakhshaliyev, Jeff Springer, Shamik Sengupta, “A system architecture of cybersecurity information exchange with privacy (cybex-p)”, in IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Accepted For Publications, 2019.

[3] Jonghoon lee, Jonghyun Kim, Ikkyun Kim, and Kijun Han, “Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles”, in IEEE Access, Accepted For Publications, 2019.

[4] Hamad Almohannadi , Irfan Awan , Jassim Al Hamar , Andrea Cullen , Jules Pagan Disso , Lorna Armitage, “Cyber Threat Intelligence from Honeypot Data Using Elasticsearch”, in IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), Accepted For Publications, 2018.

[5] Farhan Sadique , Sui Cheung , Iman Vakilinia , Shahriar Badsha , Shamik Sengupta, “Automated Structured Threat Information Expression (STIX) Document Generation with Privacy Preservation”, in 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), Accepted For Publications, 2018.

[6] Priyanka Ranade, Sudip Mittal, Anupam Joshi, Karuna Joshi, “Using Deep Neural Networks to Translate Multi-lingual Threat Intelligence”, in IEEE International Conference on Intelligence and Security Informatics (ISI), Accepted For Publications, 2018.

Thank You