A Synopsis on

AI Based SOAR

Submitted in partial fulfillment of the requirements
of the degree of

Bachelor of Engineering

in

Information Technology

by

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

Under the guidance of

Prof. Vishal Badgujar

Department of Information Technology
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI
2020-2021

CERTIFICATE

This is to certify that the project Synopsis entitled ***"AI Based Soar"*** Submitted by ***"Rahul Vast (17104042)"*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Information Technology*** to the University of Mumbai,is a bonafide work carried out during academic year 2019-2020

Prof. Vishal Badgujar
Guide

Prof. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place:A.P.Shah Institute of Technology, Thane
Date:

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

(Signature)

_____

Rahul Vast (17104042)

Date:

Abstract

Cybersecurity is becoming very crucial in the todays 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. Thus, dealing with huge variety of devices quickly will require some extent of automation in this field. Generating threat intelligence automatically and also including those which are multilingual will also add plus point to prevent well known major attacks. Here we are proposing an AI based SOAR system in which the data from various sources like firewalls, IDS, etc. is collected with individual event profiling using a deep-learning detection method. 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. For standardized format Here our system finds out about the true positive alert for which the appropriate/ needful steps will be taken such as the generation of Indicators of Compromise (IOC) report and the additional evidences with the help of Security Information and Event Management (SIEM) system. The security alerts will be notified to the security teams with the degree of threat.

Introduction

Technology is growing day by day. In this modern world, it has now become one of the integral parts of our body. It has slowly but steadily started to became everywhere around us. Even in day to day life we can see various examples of it. Our smartphones, smart watches, tabs, etc. are there beeping and glowing all around us. There are fitness trackers which keeps track of our fitness and our health conditions. There are even beds that can track sleep patterns. Automated home system that track the patterns of day to day life of the people living in it. This is just to name few. Technology has even entered in our household things and even in our home appliances. The main thing to note here is all these technologies is producing different types of data. Health related data, sensitive data, confidential data, etc. No matter what kind of data it produces it is important to protect all of these data. If not then such kind of data can actually make one vulnerable to harmful intentions of other people. That is the area where Cyber Security comes into the picture.

Cyber Security is the field which aims to protect users all around the world to protect from Cyber-attacks. This seems to be simple when we hear it first. An attack will happen and cybersecurity people will prevent it and everything will be fine. Well it's not that easy. A cybersecurity person always needs to be few steps ahead of attacker. They should be quick when deciding what to do and should not get panic themselves. Because they can often find something useful which can trace back to the attacker, it also can lead to information such as why the attack happened, how the attack happened, what were the intentions of hacker/attacker when performing such attack. Often such information on attack help us to find the answer of other questions as well such as what will be the impact of such attacks on the organization, how to avoid such attacks in future, where are we weak in terms of security and what steps needed to carry out to fill up those loop holes.

As today's world has well understood the importance of Cyber Security in their organization and there are various measures to deal with such attacks in future. There are also various tools designed by organizations to protect themselves. Such as firewalls, IPS, IDS, Antivirus, Antimalware, SIEM systems, etc. There are also practice of keeping information about the attacks happened on organization previously which is generally known as Threat Intelligence. Threat intelligence often help organizations to deal with the attacks previously happened and now they might have developed the prevention strategy for the same attack and thus attacker fails. Sharing such information with other organization can help them too to get over those attacks but it does not often happen due to privacy reason of a particular organization.

The work flow of various organizations can be different but in general all of them will have security teams divide for different tasks. Such teams often need to share their gathered information with each other and a delay in these tasks can cost a lot. There is also shortage of skilled people in this field and thus losing a potential security expert means a lot. There are often few repetitive tasks which are minor but can hold important information as well and ignoring those can cost a lot in near future. And being a human, going through such huge number of alerts and filtering out threats as false positives and true positives on daily basis can often lead to mistakes.

Thus, automating such things can help security teams to deal with more important matters which really need their attention. Automating tasks which are repetitive in nature, filtering,

classification, etc. can be done using the various new technologies and by utilizing better computation power we have now. Automatic generation of threat intelligence can also be proved beneficial to security teams. While there are progress going on taking these things into consideration now as well. We have now evolved next gen firewalls from various vendors that can carry out large amount of necessary activities which traditional firewalls can't do. Smart threat detection, Malware behaviour analyses and pattern recognition, etc. are to just name few.

Objectives

- To recognize different patterns from the event logs/dataset, detection of the vulnerabilities the loop holes in the system and prevent them.

- Event profiling of the data is to be done, so there would an ease to find out the attack, that is the type of attack would be monitored.

- The data collected from different sources must be represented into a standardized format based so there is an ease of use in classifying them.

- To translate a threat intelligence available in different languages.

- 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.

Literature Review

In literature [3], the author suggested an AI technique for Cyber-Threats Detection based on artificial neural networks. This system converts collected security events into event profiles. These profiles will then be used for cyber threat detection on a deep learning-based technique. This system is majorly based on the FCNN, CNN and LSTM neural network methods. The main goal of the system is to discriminate between true positives and false positives. This enables the security analysts to deal with significant security alerts. The results obtained after applying the model in real world suggest that it is capable of getting deployed as Network intrusion detection system. It's performance also outperforms the conventional machine learning methods. 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.

In literature [4], the author suggested a new threat intelligence technique which will be evaluated by analysing honeypot log data that will identify behaviour of attackers and by finding attack patterns. They deployed their honeypot on AWS cloud for collecting incident related data. This log data is then analysed using ELK (Elasticsearch, Logstash, Kibana) stack. These systems generate alerts and prevent Cyber Attacks based on the learned attack patterns. The potential drawback of this Model is that it requires data related to same incident in huge amount. Most of the data collected is also similar to each other which can decrease the performance of model on new data.

In literature [5], the author suggested a mechanism to represent raw cyber threat-data in Structured Threat Information Expression format in an automated manner. The method also takes care of privacy preservation. The Standard Format required for an organization to share these data with other Organizations is provided with these systems. They Improve the Privacy of The Data Because Sensitive information is removed as a CTI (Cyber Threat Intelligence) is generated. This helps the organizations to understand the threat and also make it ready for the advance analytics. This data can also be shared on a Threat Intelligence sharing platform. Thus, the system aims to bring complete automation of the security to deal with modern day attacks efficiently.

In literature [6], the authors suggested to create a neutral network that takes in threat intelligence available in different languages using which it translates that it in desired languages and thus help in making available threat intelligence in different languages. This proposed system uses Russian And English word embeddings created from cybersecurity data. It uses an LSTM based neural machine translation architecture. It also uses encoder-decoder architecture which maps Russian words to their English words. Their results show that their system easily outperforms other third-party translation engines as well as successfully detect cybersecurity terms better. The system is able to run independently in secluded operational settings. The System requirement of a cybersecurity rich data to train the model is quite high though. The system aims at making threat intelligence data available in different languages to be available globally in their natives.

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.

Proposed System Architecture/Working

SOAR stands for Security Orchestration, Automation, and Response. The main focus of this technology is to gather as much cybersecurity information as possible from various sources available and then identify the events and incidents based on the output generated by using certain algorithms. Basically, it tries to bring ease in the life of security professional by eliminating their need of constantly responding to all alerts and categorizing threats for evaluation all manually. Here the alerts are being categorized according to the degree of threat and an alert message is being sent to the user. It also tries on eliminating the false positives and detecting the true positives amongst them, playing a vital role in security orchestration.

The cybersecurity is evolving rapidly and organizations are expecting more to come from this technology of SOAR to stay ahead by rapidly responding to threat events and this making their incident response section working more efficiently. The reason for this shift could possibly be the increasing complexity of attack vectors that are evolving with the new technology coming in the market, the volume of those attacks and severity of those attacks. Satisfying the intense global regulatory demands and to fill up the gap of shortage of skilled people has also seen such automated technologies as a boon.

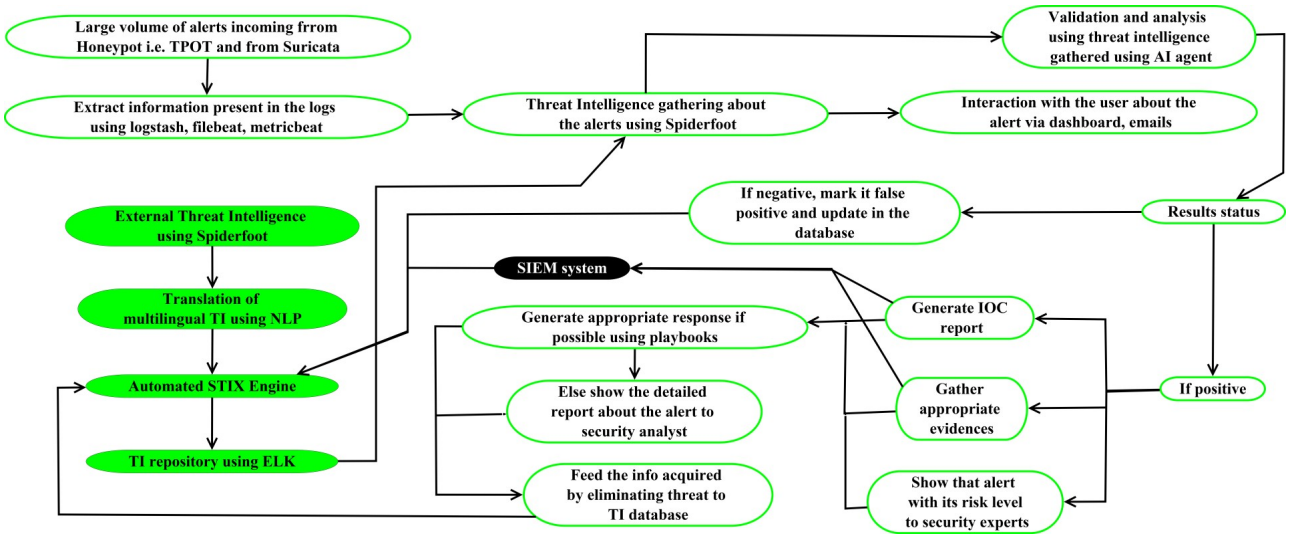

Figure 1: A.I. based Security, Orchestration, Automation and Response System Workflow

The above-mentioned existing systems tries to deal with the two important factors every organizations should have to deal with modern day attacks. That are correctly identifying true positives and making threat intelligence available to everyone without worrying about privacy. Fig. 3 is the workflow of our proposed system to deal with the modern-day threat. It is an AI based SOAR system. It tries to have the features of an existing systems along with some other features such as making available the multilingual threat intelligence and automating some repetitive tasks.

The system will have main roles of Detection of Threats, gathering intel on the threat detection, alerting incident response team about the threat with all the threat intelligence

gathered. Apart from that the system will have automated threat intelligence system along with multilingual threat intelligence also into consideration. The proposed technology work as follows: Firstly, the incoming alerts from various frontline defences like firewall, IDS, etc will be monitored by system. Information about each alert will be gathered accordingly. The system will try to get information from the Threat Intelligence System (TIS) about the alert i.e. whether there is previous known entry for the alert generated is available or not. Accordingly, appropriate steps will be taken for validating the information on alert and perform analysis operation. At the same time user whose data might get compromised will be notified or protected. Now in this entire process if the system and the user conclude that the alert is false positive. Then the entire info about the alert will be stored in TIS in appropriate format.

If the system and the user conclude that the result are true positive then the appropriate steps will be taken such as generation of Indicators of Compromise (IOC) report and gathering additional evidences with the help of Security Information and Event Management (SIEM) system. Accordingly, security teams will be notified about the alert with its threat level and other information acquired about the same.

Now if the system is able to prevent the attack then the necessary steps will be taken else will wait for the human security analyst to deal with the threat and then create appropriate record about the threat elimination from humans in Threat Intelligence System about the true positive alert.

External Threat Intelligence mentioned here is the intelligence acquired from the open source sharing platform available which will get converted into desired language set by organization and then it will get converted to standardized format by STIX Automation Engine. After that it will be finally stored in Threat Intelligence Repository. Threat Intelligence System mentioned here will be working based on the technique mentioned in [2], [5] and [6]. Based on the same methods Threat Intelligence (TI) translation and automated STIX engine will work. SIEM system mentioned is also based on [3]. SIEM's main role here is managing the information of the event, profiling it which will support STIX in its process. Apart from all there is also an phishing url detector based on [1].

Design and Implementation

Figure 2: Dashboard of our Honeypot

Above diagram is the Dashboard of our multiple honeypot built using an open source honeypot TPOT. We used this honeypot to collect the data required for training our AI model.It is deployed on AWS EC2 instance. For dashboard and storing data, ELK stack is used under TPOT Framework.

Figure 3: Attacks happened from different geographical locations

This diagram shows the attacks happened on our honeypot in counts. As we can see that most of the attacks happened from well developed and developing countries.

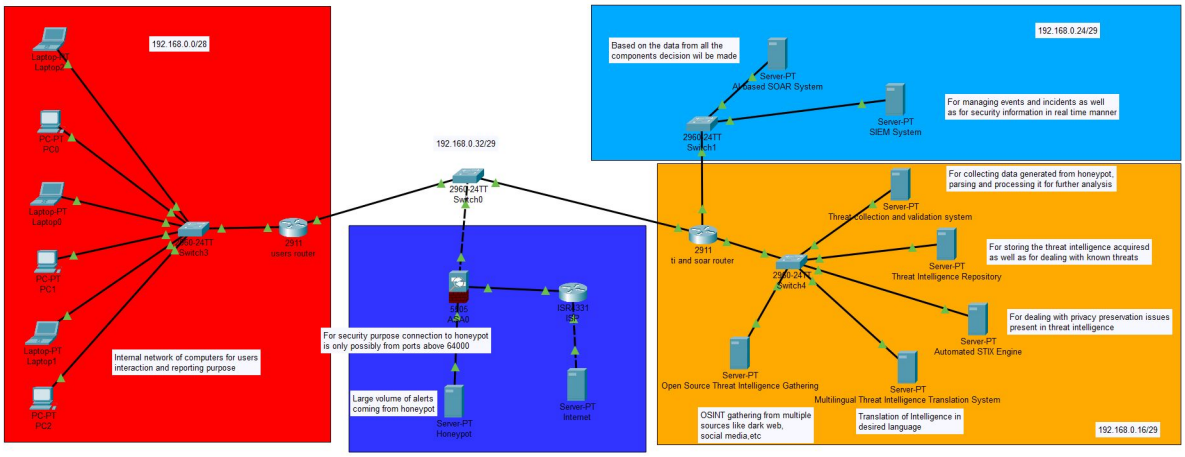

Figure 4: Planned Topology

This is our planned topology made using the Cisco Packet Tracer. It represents our general idea of how one can expect the deployment of our project after completion.

Summary

We have studied Soc operations and various tools like Soar and Siem systems as well as Threat Intelligence gathering and storing processes. Based on this we are designing an AI based system to intelligently automate these processes to bring an ease in working of the SOCs.

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.

1 Publication

Paper entitled **"Artificial Intelligence based Security Orchestration, Automation and Response System"** is accepted at **"IEEE 6th International Conference for Convergence in Technology (I2CT) 2021"** by **"Rahul Vast".**