Journal of Computer Engineering & Information Technology

Review Article A SCITECHNOL JOURNAL

Endpoint Protection of Windows Operating System using Threat Analysis Cycle

Bijaya KC* **and Roshan Chitrakar**

Department of Science and Technology, Pokhara University, Pokhara, Nepal

***Corresponding author:** Bijaya KC, Department of Science and Technology, Pokhara University, Pokhara, Nepal; E-mail: kecybijaya@gmail.com

Received date: 20 March, 2023, Manuscript No. JCEIT-23-92387;

Editor assigned date: 23 March, 2023, PreQC No. JCEIT-23-92387 (PQ);

Reviewed date: 06 April, 2023, QC No. JCEIT-23-92387;

Revised date: 24 May, 2023, 2023, Manuscript No. JCEIT-23-92387 (R);

Published date: 31 May, 2023, DOI: 10.4172/2324-9307.1000274

Abstract

This paper attempts to fill the knowledge gap in general by using machine learning within the Threat Intelligence Cycle (TIC) for proper analysis of signature based and anomaly based threat detection. This paper aims to fill the gap seen among people about proper security configuration by notifying the threat intelligence cycle and implying the significance of setting those configurations within windows 10 within DELL and HP laptops and Lenovo thinkpad within a network. Along with hardening, malicious behavior analysis is also essential to discover vulnerabilities in the private network to protect from internal threats for which the behavior analysis model is approached. For this, we have used datasets as system logs from the pfsense alert message and CICIDS2017 dataset to build a machine learning model using the xgboost classifier along with Principal Component Analysis (PCA) from which the obtained accuracy of the model is 99.75%, precision: 0.997, recall 0.998, F1 score: 0.997 for PCA 25.

Keywords: Threat intelligence cycle; Hardening windows OS; Eternal blue; Machine learning; CICIDS2017 datasets

Introduction

Securing personal systems is challenging as every system in the world is connected to a network and to the internet. Similarly, the tendency of exploiting personal data and hacking the personal system or any digital gadget is increasing in many intruders in order to get someone's personal information such that they can use it for their own benefit or the victim's loss. System hardening helps in detecting as well as closing every loophole present in the system in order to protect it from different types of cyberattacks. It involves reducing or fully cutting the pathway for attackers. For this, hardening involves segmenting resources, updating security patches, resetting default passwords and timely updating new passwords, hashing passwords, and also implementing the Principle of Least Privilege Policy (POLP). With the advancement in technology and the use of different platforms for storing data as well as information, there is an ever-increasing pace of data breaches and system vulnerabilities like DDOS, trojans, worms, viruses, and various malicious attacks. Moreover, in order to secure a whole network, it is necessary to protect every endpoint of a network so that no such issue occurs [1]. OS hardening means securing bios and base software such that a complete system and various applications in the system can be protected. Some people are still unaware of different vulnerabilities and security configurations while others don't pay proper attention to configuring and updating patches due to their busy schedules or carelessness [2]. Similarly, the lack of dedicated anomaly tracing and system configuration leads to spoofing, DOS/DDOS, and an unsafe gateway for which the vulnerable spot starts from endpoints that's why proper hardening and updating patches is a must to save self from cyber-attack and prevent several losses [3].

However, the research shows a lack of knowledge gap in people about maintaining security protocols in the endpoints. People don't know how, why, and where to configure security in the system because of which their system is compromised. Such is a problem of every organization, enterprise, finance, and government sector along with big renowned companies, and also only signature-based threat detection is insufficient nowadays due to lots of threats and vulnerabilities seen in the world [4].

To provide the solution for such problems, this paper focuses to solve the problem of cyber-attack and OS vulnerabilities in an endpoint by hardening OS, patching, signature-based, and behavior analysis along with IDS/IPS implementation in a reasonable way. The first section of our proposed methodology provides TIC for filling the knowledge gap in people or staff at enterprises and the next section is about signature and behavior analysis for detecting anomalies and abnormalities seen in the log at endpoints along with machine learning implementing IDS/IPS security threat detection and protection system which can pass alert notification. All these are done within TIC which has a plus maintenance phase finally concluding with the result and expectations [5]. This paper contributes to defining a systematic approach to fill the knowledge gap in people about security configuration by implementing the threat intelligent cycle. The paper contributes to do proper threat analysis with maintenance of security.

Literature Review

Threat is detected for finding vulnerabilities in an endpoint according to the need of network or an enterprise. System hardening, patch updating and finding threatful signature through antivirus are primary task that has been in use from ages. However, this only can't find emerging vulnerabilities in the world and people have knowledge gap like how to control over endpoints accurately so that no sensitive information is leaked and to comprehend how serious; the hazards can be [6]. This knowledge gap can be filled in people by making all the security practices systematic using threat intelligence cycle which emerged in 2013s but has not been used wisely till now. Threat Intelligence cycle has 4 phases consisting: Planning and requirement analysis, collection and processing, analysis: Signature based and anomaly based threat analysis, dissemination and feedback.

Related works

The idea of machine learning has been a good approach to analyzing behavior at the host and network especially analyzing anomalies in user behavior by authors Nassif and Nguyen. The author Nassif shows the approach to stop different cloud security threats like DDOS, IP spoofing and maintain data privacy in the cloud by using machine learning for which KDD and KDD CUP'99 dataset has been used [7]. The primary cause of data breaches especially in enterprises is due to lack of proper security configuration in endpoints in paper by chandel. End Potection Platform (EPP) and Endpoint Detection and Response (EDR) along with HIPS/HIDS has been used to analyze different features that a protection system or software should possess in order to help enterprise deal with data leakage and its avoidance. The author Chandel also highlight the importance of creating awareness on people about security configuration in endpoints as users are actually responsible for eventual security management in any enterprise. The log generated or collected by security controls have been seen to be mostly used for forensic investigation only after the consequences of the log have been detected but it is also necessary to proactively detect a breach in an enterprise for operational environment setup in an enterprise by researcher Li, et al. for this a Deep Convolutional Generative Adversarial Network (tDCGAN) has been used by the author which itself generates pseudo malware to differentiate it from real malware in paper by Kim, et al. The author Burihabwa states that for smooth OS performance, the file system should be kept safe for which the idea of SGX within a CPU has been implemented. Especially with windows OS, there are many unknown vulnerabilities, and only disabling not useful services is less efficient for which author borbor implemented a heterogeneous hardening approach to deliver some optimal solutions like firewall rule modification, disabling services, service diversification and access control. The paper by Guo is based on locating vulnerabilities in 2 parts: Approximately and accurately by data stream technique and dynamic analyzer technique. The author has used MS15-034 for this that can only check remote code vulnerability through HTTP request on windows system [8]. Similarly, in the work by Panigrahi, Ranjit, and Samarjeet Borah who have used CICIDS 2017 datasets has shown some shortcoming with the datasets for which new class labeling has been done and some classes has been merged to form new class however class imbalance problem has been found in this research work. In research by Maseer; Ziadoon Kamil; Robiah Yusof; Nazrulazhar Bahaman; and Salama A. Mostafa, and Cik Feresa Mohd Foozy states that most of the datasets used for cybersecurity are imbalance with 98% normal and 2% attacks datasets which showed large number of redundant record for which different supervised and unsupervised learning algorithm has been used. 60% of training and 40% of testing datasets have been used without proper feature selection and 5 folds cross-validation due to which the proposed model was seen time consuming like random forest in paper work. Also with the research done by Yulianto; Arif; Parman Sukarno; and Novian Anggis Suwastika where the use of Synthetic Minority Over-sampling Technique (SMOTE), Principal Component Analysis (PCA) and Ensemble Feature Selection (EFS) has been used on CICIDS 2017 datasets which gave under the Recovery Operating Characteristic curve (AUROC) of 92% and adaboost classifier using ensemble feature selection and SMOTE with accuracy 81.83%, precision 81.83%, recall 100% and F1 score 90% however the proposed methodology has not been accepted for commercial purpose and there is more chance of improving the accuracy. The author didn't propose any accurate methodology for finding vulnerability for new threats. The author Berlin Chandel and Burihabwa proposed a way of hardening OS but didn't propose any systematic way of why and in which port or at which endpoints, it is needed to do security configuration [9].

Similarly, only whitelisting and blacklisting won't be enough to protect from unknown threats. Device guard and applocker can't safeguard from fileless threats like malicious link by author Durve. Also, it is not enough to provide security configuration to administrators to protect endpoints by author Guo; Berlin; Burihabwa; and Li. The problem of updating 10,000 data using the reinforcement learning approach and system being bulky with use of EPP and EDR has been seen in the paper by Nguyen, Chandel. So, to solve this problem our paper implements behavior analysis using "splunk" which is a machine learning-based and can do users as well as entity behavior analysis to find anomalies in an effective way in any network which is implemented properly within Threat Intelligence Cycle (TIC) that define a systematic approach of configuring security practices filling the gap of awareness in people about security implementation as per the need of network [10].

Methodology

Proposed model

The proposed model consists of a number of steps such as firstly planning the safety of datasets in endpoints and then conducting all other security protocols within the threat intelligence cycle for which steps like planning, collecting, analysis, dissemination, and feedback and then the maintenance of security needs to be done (Figure 1 and Table 1) [11].

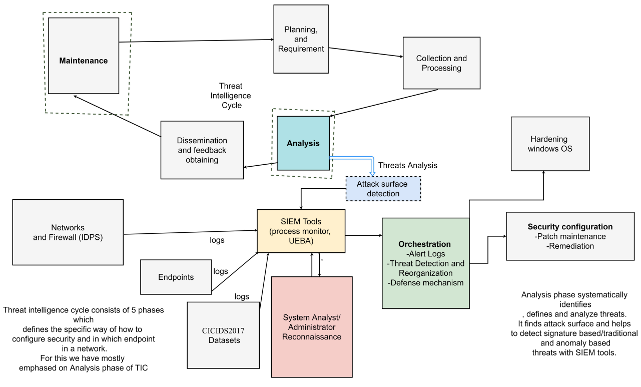

Maintenance
Planning, and Requirement
Collection and Processing
Threat Intelligence Cycle
Dissemination and feedback obtaining
Analysis
Threats Analysis
Attack surface detection
Hardening windows OS
Networks and Firewall (IDPS)
logs
SIEM Tools (process monitor, UEBA)
Orchestration
-Alert Logs
-Threat Detection and Reorganization
-Defense mechanism
Security configuration
-Patch maintenance
-Remediation
Endpoints
logs
logs
CICIDS2017 Datasets
System Analyst/ Administrator Reconnaissance
Threat intelligence cycle consists of 5 phases which defines the specific way of how to configure security and in which endpoint in a network. For this we have mostly emphased on Analysis phase of TIC
Analysis phase systematically identifies , defines and analyze threats. It finds attack surface and helps to detect signature based/traditional and anomaly based threats with SIEM tools.

Figure 1: Proposed model of the system.

S. no.	Features	Existing system	Proposed methodology
1	Manual and semi-automated windows configuration.	Yes	Yes
2	Firewall configurations.	Yes	Yes
3	Systematic security configuration with awareness on its necessity and	No	Yes

	implementation on particular endpoints.		
4	Hypervisor environment for testing.	Yes	Yes
5	Both signature and anomaly based threat detection with OS hardening using CICIDS2017 datasets,random forest algorithm, XGBoost and DDoS attack finding within TIC.	No	Yes
6	Windows server and windows security.	Yes	Yes
7	Real time data analysis by splunk of windows OS.	No	Yes
8	Threat analysis with TIC maintenance.	No	Yes
9	Digital forensic.	Yes	No

Table 1: Comparative analysis of proposed system features and existing system features.

Threat analysis model

This section includes two types of threats that can be seen in endpoints and indicates what the problem is, and how to analyze the problem (Figure 2). This consists of analyzing threat in two ways which are as below:

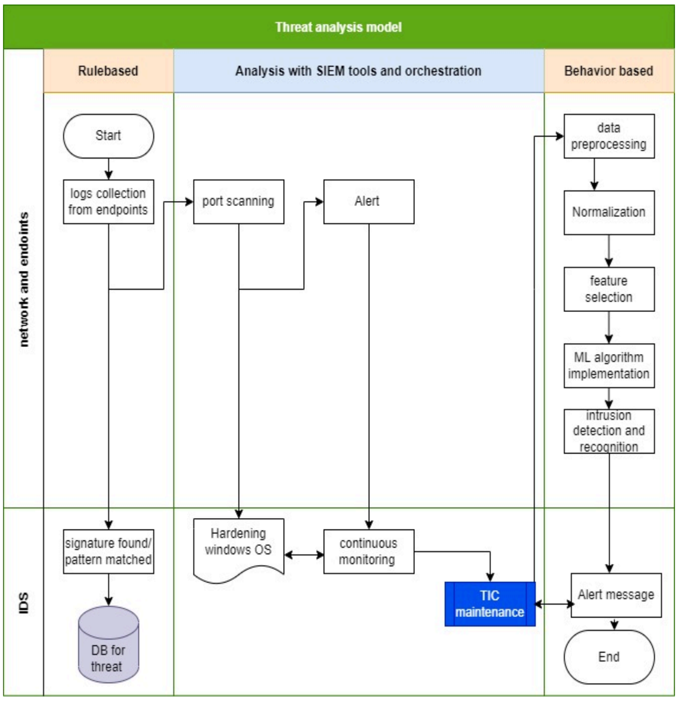


Figure 2: Threat analysis model.

Signature based threat analysis: This method is traditional one which sets certain firewalls for those malicious object like file, viruses, and so on. However, it can't protect fully from different types of malicious behavior in our endpoints mostly that unknown threats like zero-day.

Behavior based/anomaly based threat analysis: For this, we need to look over the answer to the following questions as What do users do?, How do they do?, What are their usual action?, At what location do they usually act?

For detecting behavior, both the incoming and outgoing packet from endpoints needs to be monitored and an alert is needed to be passed if some suspicious behavior is detected then signature-based as well as anomaly-based behavior analysis method is implemented.

Experiemental datasets collection

Few of the datasets have been collected by continuous monitoring of endpoints from pfsense firewall log, splunk, and other SIEM tools like process monitor and process explorer to check the overall performance of windows OS to find threatful patterns or signatures [12]. Similarly, a labeled datasets CICIDS2017 dataset has been used for training machine learning model to detect different unknown threats with proper behavior analysis. CICIDS2017 dataset comprises both normal as well as anomaly datasets with new malware attacks such as brute force FTP, brute force SSH, DoS, heartbleed, web assault, penetration, botnet, and DDoS. Name of this dataset is fundamentally based on the timestamp, source and destination IPs, protocols and attacks, source and destination ports values [13]. Computing the average of the all the words in the corpus, document vector which represent the overall category *i.e.* party tweets can be generated.

Why CICIDS2017 datasets?: The reason for this can be depicted with the comparative analysis of different datasets (Tables 2 and 3).

Comparing datasets (√=available features, ×=false or not available).

Datasets	Realistic traffic	Label data	IOT traces	Zero day attack	Full packet captured	Year
DARPA 98	✓	✓	X	X	✓	1998

KDDCUP 99	✓	✓	X	X	✓	1999
CAIDA	✓	X	X	X	X	2007
NSL-KDD	✓	✓	X	X	✓	2009
ISCX 2012	✓	✓	X	X	✓	2012
ADFA-WD	✓	✓	X	✓	✓	2014
ADFA-LD	✓	✓	X	✓	✓	2014
CICIDS2017	✓	✓	X	✓	✓	2017

Table 2: Comparative analysis of datasets.

Name of files	**Day activity**	**Attacks found**
Monday working hours.pcap_ISCX.csv	Monday	Benign (normal activities file).
Tuesday working hours.pcap ISCX.csv	Tuesday	Benign, FTP-patator, SSH-patator.
Wednesday working hours.pcap_ISCX.csv	Wednesday	Benign, dos goldeneye, dos hulk, dos slowloris, dos slowhttptest, heartbleed.
Thursday-workinghours webattacks.pcap_ISCX.csv morning	Thursday	Benign, web attack brute force, web attack sql Injection, web attack-XSS.
Thursday-working hours afternoon infilteration.pcap_ISCX.csv	Thursday	Benign, infiltration.
Friday-workinghours morning.pcap_ISCX.csv	Friday	Benign, bot.
Friday-working hours afternoon DDos.pcap_ISCX.csv	Friday	Benign, portscan.
Friday-workinghours afternoon PortScan.pcap ISCX.csv	Friday	Benign, DDoS.

Table 3: Describing CICIDS2017 datasets.

CVSS rating criteria

CVSS score can offer assistance to distinguish, track, and remediate any vulnerabilities that debilitate your commerce and the arrangement. This system makes a difference to guarantee privacy with respects to their user's delicate information to organize data for businesses and maintain security in a network. The base score breakdown includes:

0.0=No risk to the system
0.1-3.9=Low
4.0-6.8=Medium
7.0-8.9=High
9.0-10.0=Critical.

Results and Discussion

This section contains expected outcomes and finding of our research. Evaluation measures used are accuracy, precision, recall and F1-score for a machine learning model like as for finding signature based threats evaluating SIEM tools accuracy to find threats CVSS scoring is done (Table 4).

Tools used: Some of the tools used for the proposed methodology are:

Asset ID	Asset name	Asset classification	Model	Types of assets	Purpose
AST-100	Windows server	Operating system	Windows server 2012 R2	Software	Host applications.
AST-101	Mikrotik router	Router	Mikrotik hAP ac	Hardware	Internet connection.
AST-102	Mikrotik switch	Switch	Mikrotik CRS112-8G-4S-IN	Hardware	Intranet connection.
AST-103	Laptops	Laptop	HP elitebook	Hardware	Deployment and testing.
AST-104	Laptops	Laptop	HP Envy	Hardware	Development activities.
AST-105	Laptops	Laptop	Lenovo thinkpad	Hardware	Deployment and testing.
AST-106	pfSense	Firewall	FreeBSD	Software	Filter the network, prevent unauthorised network access.
AST-107	VMware	Application	VMware workstation 16 pla	Software	Virutalisation.
AST-108	Cisco packet tracer student	Application	6.2.0	Software	Routing and Switching.
AST-109	Kali GNU/linux rolling	OS	Linux 5.16.0-kali3-amd64	Software	Penetration testing.
AST-110	Windows 10	OS	Home 64-bit (10.0, build 19	Software	GUI based OS for IOT devices.
AST-111	Wireshark	Application	3.6.1	Software	Network protocol analyzer.
AST-112	Metasploitable framework	Framework	MSF6	Framework	Modular penetration testing platform in kali linux.
AST-113	Anaconda	Application	Anaconda.Navigator 3	Software	For building machine learning model and its prediction.
AST-114	Scikit-learn	Libraries		Tools	Data analysis and Implement machine learning.
AST-115	Numpy	Libraries		Tools	For working with array and list. Code optimization.
AST-116	Pandas	Libraries		Tools	Statsistcal analysis, data normalization and visualization.
AST-117	Notebook	Application	6.4.5	Software	For python programming and machine learning model building
AST-118	XGBoost	Algorithm		Classifier	For classification.
AST-119	Python	Programming languag	3.10 (64-bit)	Software	For machine learning.
AST-120	Process explorer	Application	16.43	Software	For task and system monitoring.
AST-121	Process monitor	Application	3.89	Software	Advance monitoring tool for windows OS.
AST-122	Splunk enterprise	Application	8.0.4	Software	To search, analyse and visualize detail information of endpoint.
AST-123	Zenmap	Application	7.92	Software	Network exploration and security auditing.

Table 4: Tools and device used.

The experimental environment

This section deals with the implementation and analysis of the endpoint protection of windows OS by behavior analysis within TIC [14]. The experiment was performed in Jupyter notebook along with SIEM tools with windows 10 home 64-bit (10.0, build 19043) HP ENVY 15 × 360 PC, bios F.11, 8192 MB RAM and 2.6 GHZ Intel core 4510u CPU processor while linux has been used as an attacking source to the endpoint (Figure 3).

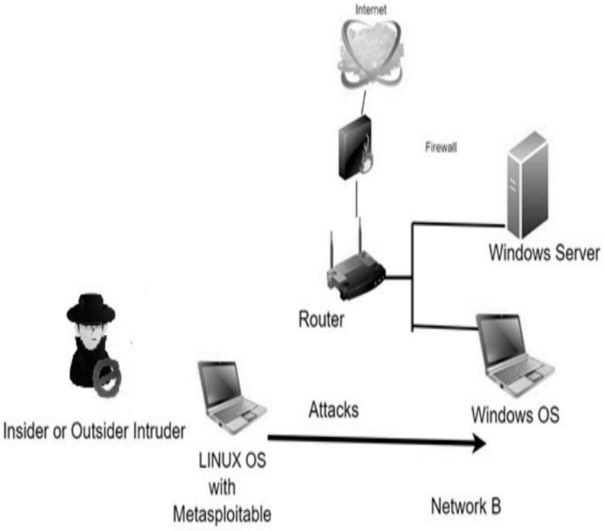


Figure 3: Experimental network prototype.

Implementation scenario

The intelligence cycle is represented where each phase is carried simultaneously to analyze data and produce meaningful intelligence in order to maintain quality of security features. The transformation of large source of data that has been gathered from authenticated endpoints in network is done after proper analysis into suitable form for data normalization, translation, decryption, protection and storage along with decision making through statistical procedure (Figure 4).

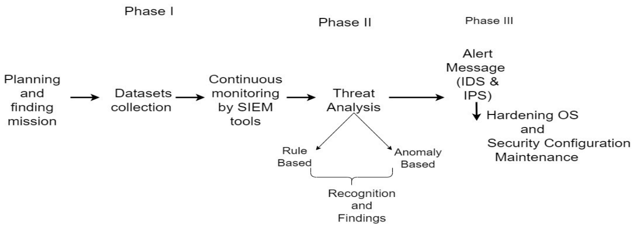

Phase I
Phase II
Phase III
Planning and finding mission
Datasets collection
Continuous monitoring by SIEM tools
Threat Analysis
Alert Message (IDS & IPS)
Hardening OS and Security Configuration Maintenance
Rule Based
Anomaly Based
Recognition and Findings

Figure 4: System pipeline for endpoint protection.

First of all, planning and finding mission has been done to protect endpoints and continuous monitoring has been done by SIEM Tools to collect datasets, find threats and harmful patterns in any link/file or incoming and outgoing traffic. Then rule based threats has been found using SIEM tools and for finding more unknown vulnerable threat labeled datasets CICIDS2017 has been used for building, training a model and then predicting threats. As per the pipeline, the phases [15].

First phase consists:

- Planning and finding mission.
- Dataset collection from endpoints.
- Continuous monitoring by SIEM teams.

Second phase consists of

- Threat analysis
- Rule based/signature based threat analysis.
- Anomalies based threat analysis.
- Recognition and detecting.
- Alert message.

Third phase consists of

- Orchestration.
- Hardening OS.
- Maintenance.

Conducting phase I we find that

- Windows OS should be protected with firewall setup at the network for inbound traffic.
- Proper monitoring of endpoints with windows OS and router in which it is connected should be done by SIEM tools.
- Data collection or log record collection should be done for analysis.
- Some malicious file could be auto detected by windows defender and deleted while some of those were still undetectable to windows defender (Tables 5-7).

Malware analysis: Wireshark has been used.

Title	Malware traffic detection
Tool	Wireshark
Criticality	High
Description	Captured Network traffic and filtered Http and TCP logs to check and verify host and destination source Id and message passed by Wireshark.

Process of analysis	Download log file and check to verify it.
Curative	Use active log analysis in Wireshark and then use IDS/IPS.

Table 5: Malware analysis by wireshark.

Windows events logs analysis by process monitor

Title	**Windows OS endpoints analysis by process monitor**
Tool	Process monitor.
Criticality	Medium.
Process of analysis	Downloaded process monitor and by running it all events collected.
Curative measures	Detecting threat using virustotal.com and using IPS and hardening Windows OS.

Table 6: Windows OS analysis by process monitor.

Windows event log analysis by process explorer

Title	**Windows event analysis and browsing history detection by process explorer**
Tool	Process explorer.
Criticality	Low
Process of analysis	Downloaded process explorer and run it to check windows log events, processor used, memory consumption and browsing history. Used exebinder.exe to check threat in virustotal.com
Curative measures	Windows defender, IDS and IPS implementation and killing vulnerable process.

Table 7: Windows event log analysis by process explorer.

Phase II: Signature based intrusion analysis workflow

- Collected data packets monitoring.
- Data processing and detecting threats using snort suricata and firewall.
- If packet matches, pass alert log and then do hardening of windows OS, port scanning, use antivirus else discard the packet (Figure 5).

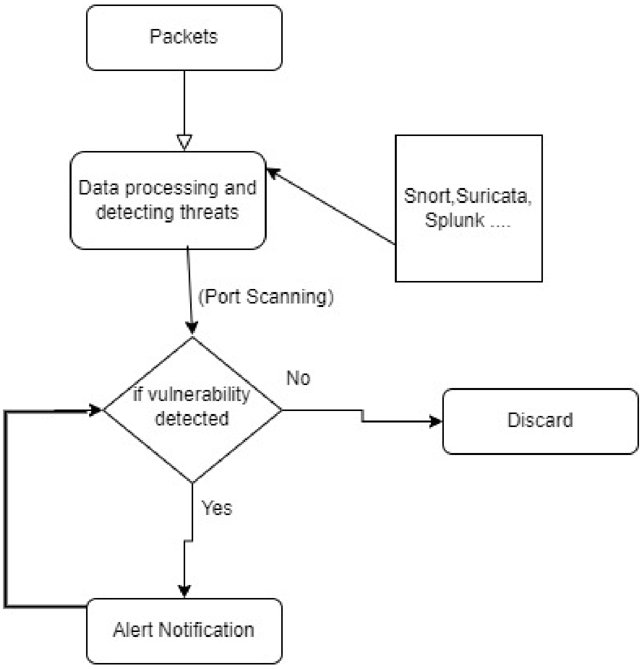

Packets
Data processing and detecting threats
Snort,Suricata, Splunk ....
(Port Scanning)
if vulnerability detected
No
Discard
Yes
Alert Notification

Figure 5: Packet analysis workflow to find signature based threat.

Packet monitoring and packet analysis as well as data processing should be done. For detecting any threat Alert notification should be set up in an endpoint which can also be done automatically by SIEM tool's orchestrations that can easily find rule-based threats but the machine learning tool like splunk can detect far more abnormalities as well as predict the system performance (Figures 6 and 7, Tables 8-11).

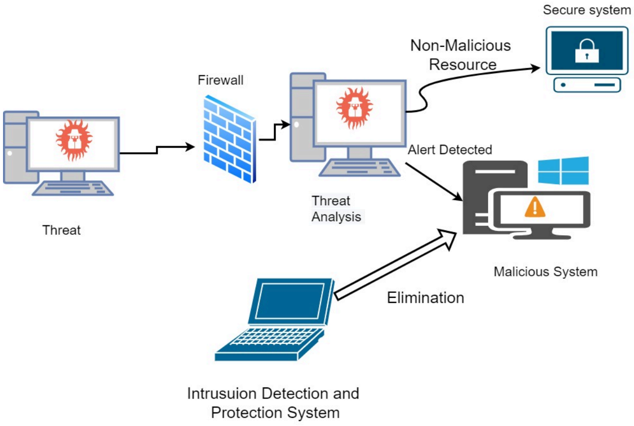


Figure 6: Threat addressing process.

Tool	Snort
Detection	Defines malicious activity over the network and also uses those rules to find packets that match against them and generates alerts for users.
Control measures	Act as a packet sniffer and also as a packet logger. This is useful for traffic debugging in the network.

Table 8: Snort analysis and control measure details.

Tool	Suricata
Detection	Suricata does deep level packet inspection, pattern matching that makes it incredibly useful for attack and thread detection.
Control measures	It supports hashing, files extraction and it has hooks for the lual scripting language, which can be used to modify outputs and even create complex and detailed signature detection logic.

Table 9: Suricata analysis and control measures details.

Endpoint	Windows server 2008 R2 with service pack 1 with security update rollup March 19, 2019
Title	Exploit loophole of eternalblue on windows server with metasploitable.
Criticality	High
Description	EternalBlue, known as MS17-010, is vulnerability in microsoft's Server Message Block (SMB) protocol. This is exploited by using metasploitable framework.

Threat measures	First we search eternalblue to exploit and use exploit/windows/smb/ms17 010 eternalblue after that we set payload, rhosts, lhosts and rport then execute the command to exploit the system as shown in below figures. To verify the compromised target, running commands such as sysinfo to obtain OS information. Also using shell to verify ipconfig of the compromised host and getuid to get the current username.
Curative measures	Later security patch were updated by microsoft and solved.

Table 10: Exploiting eternalblue on windows server with metasploitable.

Tool	**Splunk**
Detection	It identifies data packets, find problems in an entity and provide metrics as well as intelligence in analyzing threats.
Control measures	Worms and vulnerability detection as well as user and entity behavior analysis in a particular system of remote, local device and overall machine data.

Table 11: Splunk detecting threats and its curative measures.

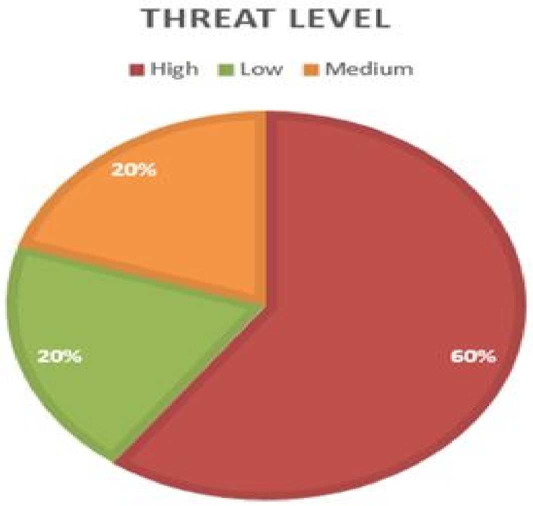


Figure 7: CVSS Scoring and threat level found from signature-based threat detection.

Anomaly/behavior analysis workflow

With proper analysis of what the user is doing? from what location? Is the behavior threatening or not? and if our windows OS is safe or not? Is the activity done in windows OS endpoints justifiable or not? we need to know about all these with proper observation of user behavior. for this we need to know about someone’s location, their IP address, and if the user is authorized or not. All these question helps in finding a behavioral pattern to detect the threat and find abnormalities (Table 12).

Endpoints	**Tor and proxychains**
Title	Anonymous account login using tor and proxychains.
Criticality	High
Description	Anonymous login defines user login without using validated or authorized user name and password so that user can't be traced.

Threat measures	Can scan the vulnerable ports by using proxy servers.
Curative measures	Firewall enabling, IDS and IPS alert system, windows server hardening.

Table 12: Anonymous login from linux to attack windows server.

The algorithm for anomalies behavior analysis consists of following process:

- CICIDS2017 dataset collection.
- Data pre-processing.
- Normalization, standardization and feature selection.
- Split the datasets into train and test datasets.
- Choose classifier.
- Train the classifier.
- Find performance criteria and Do model evaluation.
- Tune the hyper parameter.
- Model prediction (Figure 8).

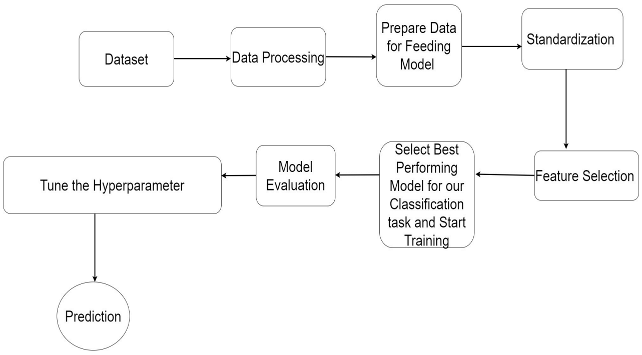

Dataset
Data Processing
Prepare Data for Feeding Model
Standardization
Feature Selection
Select Best Performing Model for our Classification task and Start Training
Model Evaluation
Tune the Hyperparameter
Prediction

Figure 8: Implementation scenario: ML model for behavior based threat detection.

Data pre-processing

It is the process of transferring raw data into meaningful and valuable datasets. In Machine Learning (ML) forms, data preprocessing is basic for guaranteeing huge datasets are designed in such a way that the information they contain can be translated and parsed by learning calculations.

Steps in data-preprocessing are:

- Import the libraries.
- Import the data-set.
- Check out the missing values.
- See the categorical values.

Prepare data for feeding model

We do proper sampling and calculate standard deviation before feeding datasets into the model. Standard deviation is the average deviation of every value from the mean value in given datasets.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{X})^\wedge 2}{n}}$$

Where denotes the standard deviation. (N/B: This is lower case sigma).

x denotes each individual value in the data set.

x̄ denotes the sample mean. n is the number of values in the sample data set.

Feature selection and extraction

When there's a feature set F={f1, fi,…, fn} the issue in feature choice is to discover a subset that differentiates patterns whereas maximizing the learner algorithm’s execution capacities. Thus, the list of feature extraction algorithms’ scoring work is indicated by F’, the subset to be found. For optimality in feature extraction in machine learning, the feature is fetched around finding the scoring feature’s expanding include or ideal feature or just removing the unwanted features and taking the rest.

Create the train and test datasets

Machine learning model learns from your information to create expectations and find more accurate findings. So the data-set is generally divided into 80:20 ratios, where 80 percent is to train the model and 20 percent is used to test the model accuracy.

Standardization

Data standardization rescales the attributes so that the mean is 0 and the variance is 1 or in any range. Standard scaler is utilized to the dataset straightforwardly to standardize the input variables. To begin with, a standard scaler occurrence is characterized with default hyper parameters.

Choose PCA

The principal component examination could be a well-known unsupervised learning procedure for diminishing the dimensionality of datasets. It could be a procedure to draw solid patterns from the given dataset by diminishing the fluctuations.

Choosing number of components in a PCA using explained variance method

In the underneath plot, each bar appears the explained change rate of individual components, and the step plot appears the cumulative explained change rates. By looking at this plot, we are able effortlessly to choose how numerous components ought to be kept. In this illustration, only the 25 components capture nearly all the fluctuation within the dataset. So, we choose to choose as it were 25 components (Figures 9 and 10).

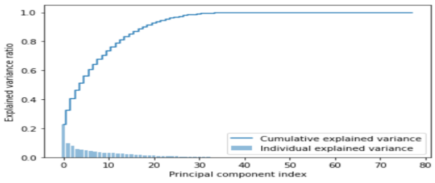


Figure 9: Histogram showing variance ratio and principal component index without using n_components.

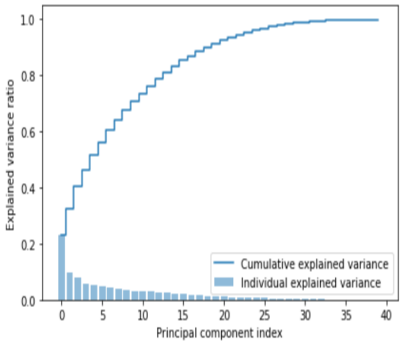


Figure 10: Histogram showing variance ratio and principal component index when used n_components.

Choose classifier

A random forest may be a machine learning method that's utilized to unravel regression and classification problems. It utilizes ensemble learning, which uses strategy that combines numerous classifiers to solve complex problems.

However, with our datasets we find more accuracy with XGBoost so, it has been used as final classifier. XGBoost is an execution of gradient boosted decision trees outlined for speed and execution. XGBoost could be a versatile and profoundly exact execution of gradient boosting that pushes the limitation of computing control for boosted tree algorithms, being built to a great extent for energizing machine learning model execution and computational speed.

Train the classifier

Training the classifier is done from the train datasets formed from the above procedure. Firstly, we train using random forest classifier for validating our experiment with the previous task, and then we use XGBoost to train it finally for better accuracy.

Find the performance criteria

The issue of predictive modeling is to form models that have great execution making predictions on modern unknown datasets. The execution of your machine learning models is done to know: How to examine your model utilizing the training dataset?, How to assess your model employing a random train and test split?

Model evaluation

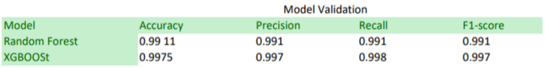
Machine learning model evaluation is done with.

Confusion matrix

Confusion matrix is a matrix that depict the accuracy of a model on a set of dataset or ground label dataset for which the labeled true value has been already defined.

The confusion matrix has following terms

- In TP, the model predicted correct and it was actually correct.
- In TN, the model predicted correct but actually it was the not correct, also called false alarm.
- In FP, the model the false but actually it was true/correct.
- In FN, the model predicted correct but actually it as incorrect (Figure 11).



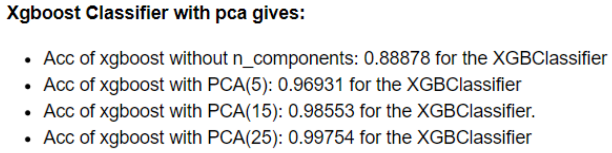
Model Validation

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.99 11	0.991	0.991	0.991
XGBOOSt	0.9975	0.997	0.998	0.997

Figure 11: Model validation with confusion matrix.

Model prediction

Model prediction helps in knowing whether the build model can accurately predict the datasets or not. The result of our experiment shows our model can predict with 99.75% accuracy (Figure 12).



Xgboost Classifier with pca gives:

- Acc of xgboost without n_components: 0.88878 for the XGBClassifier
- Acc of xgboost with PCA(5): 0.96931 for the XGBClassifier
- Acc of xgboost with PCA(15): 0.98553 for the XGBClassifier.
- Acc of xgboost with PCA(25): 0.99754 for the XGBClassifier

Figure 12: Findings with a chosen classifier XGboost.

Comparative analysis

In this research work, we will be comparing previously done work with our proposed methodology (Tables 13 and 14).

Thesis work/paper	SIEM tools	Remarks
2	EPP md EDR	Delay response and might cause system to slow down.
4	Device hard and applocker whitelisting	Version incompatibility with all windows and lurrdware requirement might not match.
3	FHewall, IDS and IPS	Only hardening.
6	Wireshark	Network traffic log monitoring.
Proposed method	Splunk, process explorer and process monitor. WH hark, pfSense.	Both user and ent'lty monitoring, econorrncal and easy continuation.

Table 13: Comparative analysis of SIEM tools used.

Thesis work/papa	Algorithm used	Accuracy	Precision	Recall	F1 score	Finding
21	New class label were formed by merging and splitting class with their prevalence rate applying simple techniques.					Class imbalance problem seen. 83.34% for majority class and 0.00039% for minority class (heartbleed). No proper work is done.
22	ANN, SVM, decision tree, naive bayes, random forest, k-means, self organization map. Expectation maximization.	With random forest accuracy is 99.54%	With Random forest precision is 99.56%	With random forest recall is 99.55%	With random forest F1 score is 99.55%	With random forest 9.38sec training time and 6.76 sec test time. this detail is obtained from dataset of the ISCX consortium using 40%-60%, 50%-50%, or 60%-40% train and test datasets respectively.
23	SMOTE, PCA. EFS	81.83%	81.83%	100%	90.01%.	This detail is obtained from mondays working hour dataset using 70% train and 30% test datasets.
Proposed work	Random forest with PCA, using n_components= 25.	Acc. of random forest with PCA 5: 0.95780, Acc. of random forest with PCA 15: 0.98835, Accuracy of random forest with PCA 25: 0.99113 for the random forest classifier modelaccuracy: 0.9911.	Precisio: 0.991	Recall: 0.991	F1 Score: 0.991	More accurate than previous model.
Proposed work	XGBoost with PCA using n_components=25 (final one)	Acc. of XGBoost with PCA (5): 0.96931, Acc. of XGBoost with PCA (15): 0.98553, Acc. of XGBoost with PCA (25): 0.99747 Model accuracy: 0.9975.	XGBoost precisio: 0.997 for PCA 25	XGBoost recall: 0.998 for PCA 25	XGBoost F1 score: 0.997 for PCA 25	Accuracy enhanced with this than with random forest.

Table 14: Comparative analysis of paper work with CICIDS 2017 datasets for anomaly detection.

Phase III

After completion of connection and address table, here first we have to check router version and if it's needed then security technology package should be enabled on R1. In the initial step ping from PC-1 to PC-0 is successful. Also ping from PC-0 to PC-1 is successful.

Also creating IOS IPS configuration directory in flash and also creating IPS rule and signature storage location in R1. Now I have to enable syslog if it is not enabled and also need to enable the timestamp service to complete the configuration part (Figure 13 and Table 15).

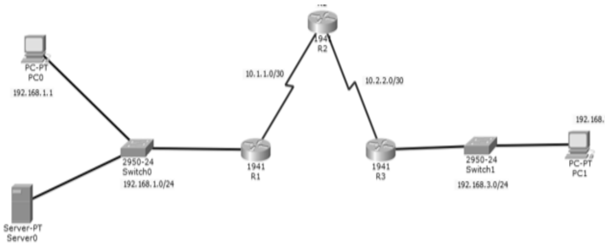


Figure 13: Simulation of mine proposed system in cisco packet tracer.

End point	Cisco packet tracer
Objectives	Enable IOS IPS. IPS signature. Syslog IPS configuration. Verify IPS.

Table 15: Objectives of cisco packet tracer.

OS hardening

Operating system enhancements typically include:

- Follow security best practices to ensure a secure configuration.
- Automatic operating system updates with patches and service packs.
- It provides additional security measures such as firewalls, endpoint protection systems, and operating system security extensions.

Patch management: Patch management is the strategy for recognizing, procuring, introducing, and confirming program patches for the system frameworks and its core parts, and products. Security patches rectify useful issues in the program and firmware. Patches serve on numerous other purposes than fair settling program imperfections; they can moreover include unused highlights to computer program applications and firmware form, counting security capabilities. To solve the problem of eternalblue, we need to do updating of windows OS.

Group policy for hardening: Using command gpedit on the command line, you can open a group policy that allows configuring access protocols, and software configuration to allow or disallow to run in the system for different user and admin accounts as well.

Windows powershell management: Powershell allows maintaining various security configurations in our endpoint like enabling VPN with command Vpnconnectiontriggerapplicationname"<Vpnconnection>"–ApplicationID"<AppPath>".

Windows defender activation all time: It provides different security features to our endpoint which has already been given by microsoft so, with windows defender we can secure our system from different vulnerabilities. It is simple, easy to use and provides a middle level security protection and can be a good security product for home users. Microsoft defender has been included in each form of windows since vista, and is presently fair one of a set of built-in free security instruments.

Maintenance

Maintaining security is a must in order to protect system for long run and focus on core and prioritized aspects of an organizational need or a system need to protect from any kind of threats. Using all the maintenance procedure mentioned below threat level can be minimized like threat with high CVSS scoring and finally solved. From all the experiment done and result found we see the need and importance of maintenance. Maintenance can be divided into different phases like: External and internal monitoring, planning and source verification, vulnerability assessment and threat analysis, timely scanning and patch review, security orchestration and remediation (Figure 14).

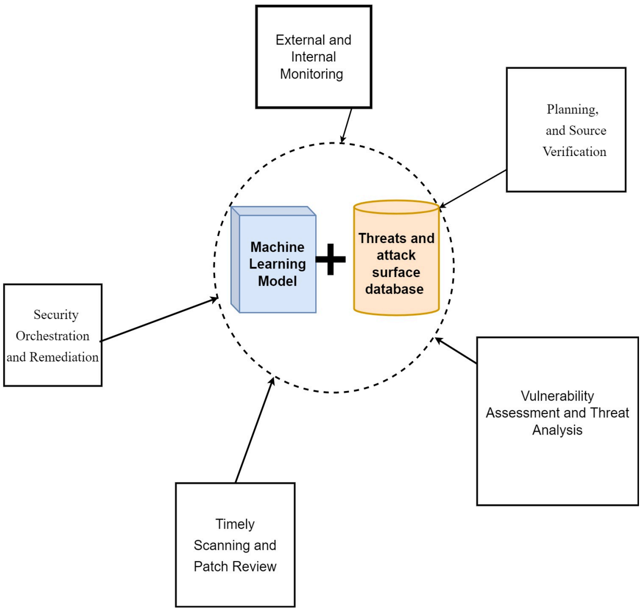

External and Internal Monitoring
Planning, and Source Verification
Machine Learning Model
+
Threats and attack surface database
Security Orchestration and Remediation
Vulnerability Assessment and Threat Analysis
Timely Scanning and Patch Review

Figure 14: Security maintenance model.

External and internal monitoring: Monitoring the external environment is all about examining threat, vulnerabilities, incoming traffic and all datasets needed to make a decision like we have collected logs from wireshark and pfsense to find threats and vulnerabilities.

Internal monitoring includes monitoring endpoint, its processing, and accessed products as well as incoming traffic and outgoing traffic like with splunk we have monitored user and entity analysis to protect our endpoints. Setting abnormal login attempt alert and DDoS prevention has been done with splunk. Monitoring Network devices, and channel as well as doing real-time analysis of windows OS has been done and all security has been monitored continuously for protecting endpoints. Thus, both external and internal monitoring helps in finding whether our endpoint is safe or not.

Planning and source verification: Planning and source verification is a must to do risk assessments which includes looking over entire system in order to reduce further risk and prioritizing risk should be done as per their severity and proper planning should be done to protect whole system from vulnerabilities. Like we have firstly planned to protect our windows OS endpoints and checked about all the incoming and outgoing traffic as well as access details. All the source must be verified for protecting system from anonymous access and any type of unknown risks.

Vulnerability assessment and threat analysis: Vulnerability assessment and threat analysis should be done like finding a vulnerability and protecting it with IDS/IPS has been done in this paperwork detects malware and other vulnerabilities and provides alert log as well as block malware files, traffic. Here we have done CVSS scoring to determine the severity of Risk. Threat using attack database and using machine learning model that detects and predicts the Threat occurred.

Timely scanning and patch review: Timely scanning and patch review can be done which helps in maintaining security protocols every time an endpoint is used. From time to time updating, maintaining security credentials, and disclosing any loopholes after passing the alert message using splunk and IPS have been done here with our paper work. Patches for eternalblue have been applied.

Security orchestration and remediation: Security orchestration and remediation is one of the milestone for security maintenance as with this every SIEM tools work effectively to continuously maintain security and monitor endpoints like the process monitor, process explorer and the splunk used. Also, the machine learning model that has been built can be deployed further for security orchestration. We have suggested and used different curative measures for different types of threats like hardening OS, patch updating, using antivirus as well.

Conclusion

The maintenance framework is considered as imperative components to develop an arrangement of conspire for compelling patching and convenient upgrading in end devices along with proper analysis of every sort of threats. Introducing, timely monitoring and upgrading security systems and patches can offer assistance organization gadgets to self-concise the issue and alarm if it's useful. Although different threats are known still many threats are undetectable in this world so, for this we have built a model using random tree classifier using PCA with n_components 5, 15, 25 and XGBoost using PCA with n_components 5, 15, 25. Accuracy of random forest with PCA (5): 0.94892, PCA (15): 0.97634, PCA (25): 0.99118 for the random forest classifier while accuracy of XGBoost with PCA (5): 0.96931, XGBoost with PCA (15): 0.98553, for the XGBoost classifier so from this we find more accuracy with XGBoost and we get maximum scores by using n_features=25 which exceeds out threshold accuracy 85%. Similarly, model accuracy: 99.12% for PCA 25 and precision: 0.991, recall: 0.991, F1 score: 0.991 for random classifier with PCA 25 Accuracy of XGBoost: 0.99747 for the XGBoost classifier, model accuracy: 0.99754, precision: 0.997, recall: 0.998, F1 score: 0.997 for XGBoost classifier.

Thus, final result of this experiment gives a more accurate model for detecting new and unknown threat with our XGBoost model while all other known threats has been uncovered with signature-based model for all security related issues from all the abnormalities, inconsistent datasets and handling missing values and improving accuracy ratio.

Future Enhancements

The SIEM tool like splunk used for orchestration and security issue finding can be made more systematic to find threats with updated features in future as it has more potential rather than just finding user and behavior analysis for protecting endpoints and as well as finding remote data having threat sources. Also lots of work can be done to enhance maintenance phase like compliance editing, vulnerability assessment and risk assessment along with risk mitigation plan as backup. When the number of datasets increases then it increases model accuracy however it also brings more load on server and this can take more time and space along with good GPU supported system for proper prediction else can lead to information loss. So, some filtering methods can be used to limit the missing values. Also this model can be deployed at cloud to find more security issues but it will take some extra time and mathematical technique as well. Some other methods like Artificial Neural Network (ANN), convolution neural network can also be used for further work.

References

1. Tan MK, Goode S, Richardson A (2021) Understanding negotiated anti-malware interruption effects on user decision quality in endpoint security. Behav Inf Technol 40:903-932.
2. Chandel S, Yu S, Yitian T, Zhili Z, Yusheng H (2019) Endpoint protection: Measuring the effectiveness of remediation technologies and methodologies for insider threat. 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC) 81-89.
3. Borbor D, Wang L, Jajodia S, Singhal A (2017) Securing networks against unpatchable and unknown vulnerabilities using heterogeneous hardening options. IFIP Annual Conference on Data and Applications Security and Privacy 509-528.
4. Waterson D (2020) Managing endpoints, the weakest link in the security chain. Netw Secur 2020:9-13.
5. Chen J, Liu W, Lv X, Ji D, Shi J, et al. (2019) Research on microkernel-based power dedicated secure operating system. J Signal Process Syst 91:1127-1136.
6. Forte D (2006) Endpoint and perimeter security: A new symbiosis. Netw Secur 2006:7-8.
7. Hwang C, Hwang J, Kwak J, Lee T (2020) Platform-independent malware analysis applicable to windows and linux environments. Electronics 9:793.
8. Rizvi SK, Aslam W, Shahzad M, Saleem S, Fraz MM (2022) PROUD-MAL: Static analysis-based progressive framework for deep unsupervised malware classification of windows portable executable. Complex Intell Syst 1-3.
9. Rhode M, Burnap P, Jones K (2018) Early-stage malware prediction using recurrent neural networks. Comput Secur 77:578-594.
10. Karantzas G, Patsakis C (2021) An empirical assessment of endpoint detection and response systems against advanced persistent threats attack vectors. J Cybersecur Priv 1:387-421.
11. Haddad J, Pitropakis N, Chrysoulas C, Lemoudden M, Buchanan WJ (2023) Attacking windows hello for business: Is it what we were promised?. Cryptography 7:9.
12. Case A, Jalalzai MM, Firoz-Ul-Amin M, Maggio RD, Ali-Gombe A, et al. (2019) HookTracer: A system for automated and accessible API hooks analysis. Digit Investig 29:S104-S112.
13. Hull G, John H, Arief B (2019) Ransomware deployment methods and analysis: Views from a predictive model and human responses. Crime Sci 8:1-22.
14. Shoffner M, Owen P, Mostafa J, Lamm B, Wang X, et al. (2013) The secure medical research workspace: An IT infrastructure to enable
6:222-225.
15. Lewis N, Case A, Ali-Gombe A, Richard III GG (2018) Memory forensics and the windows subsystem for linux. Digit Investig 26:S3- S 11.