An in-depth investigation of cyber-attack and cyber security: emerging trends and current events

Chinmay Anand
dept. Computer Science and Engineering
New Horizon College of Engineering (VTU)
Bengaluru, India
1nh19cs037.chinmayanand@gmail.com

Shreya Korada
dept. name of organization
New Horizon College of Engineering (VTU)
Bengaluru, India
1nh19cs167.shreyakorada@gmail.com

Raksha S
dept. name of organization
New Horizon College of Engineering (VTU)
Bengaluru, India
1nh19cs140.raksha.s@gmail.com

Abstract— **The intricacy and originality of harmful software (also known as malware) have grown with time, despite the fact that it has existed since the invention of computers. The term "malware" (or "malicious code") refers to software that carries out an attacker's malicious intent. Finding out how a piece of malware behaves and what it's likely to be used for is called malware analysis (such as Trojans horse, worms and virus). Malware may be identified and categorized into recognized families by exploiting its behavior and any patterns that follow it, whether they are static or dynamic. This survey study will give an outline of the methods for examining and categorizing the malware.**

Keywords— Cyber Threat, malware, Antivirus, cyber security, deep learning, malware analysis, malware detection

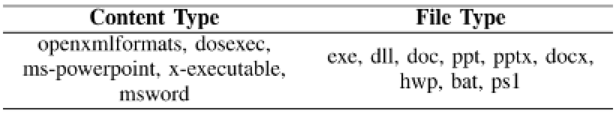
I. INTRODUCTION

Malware is the most dreadful and serious security problem that the Internet faces today. The Internet, which has become an integral part of our lives by allowing us to communicate with people all over the globe, has also been a big setback by offering up multiple routes for those with malign intentions who seek to target and injure legitimate users in various ways for various causes. In order to propagate malware, attackers target vulnerabilities in browsers, web services, and operating systems, or they utilize social engineering tactics to trick people into running harmful code. In order to avoid detection by standard barriers like firewalls, antivirus software, and gateways—which often utilize signature-based approaches and are unable to identify the previously unknown harmful executables—malware authors employ a variety of obfuscation techniques such instruction replacement, dead code insertion, register reassignment, subroutine reordering, code transposition, and code integration. Since they must first examine zeroday malware in order to establish their signatures, commercial antivirus companies are unable to provide rapid protection against it. In this article, malwares are briefly discussed along with an overview of the many types of malwares, the evolution of malware camouflage, malware obfuscation strategies, malware analysis techniques, and malware detection methods. Additionally, a comparison of analysis and detection methods has been done, and emerging patterns in the field of detection methods have also been identified.

II. LITERATURE SURVEY

A. Analysis of Today's Malware and Its Distribution Network

Malware is rogue software that is developed to do unlawful activities such as inflicting malfunctions, exposing a computer to attack, destroying software, and so on. [1] Malware has evolved into one of the subjects on which the academics have been focusing as a result of these intrusive actions occurring on a global scale. The possibility of identifying malware through clustering at the pre-distribution stage by manually analyzing samples that are unknown in the cluster has been examined in connection to the detection of malware through binary clustering. Clustering malware enables academics and industry to have a better understanding of scattered malware. It keeps track of variations and offers a reliable way for classifying and categorizing unknown malware samples. The clustering findings highlight various existent variations, such as their bytes level properties, as figure [1], which are easily detected using a similarity hash.



Content Type	File Type
openxmlformats, dosexec, ms-powerpoint, x-executable, msword	exe, dll, doc, ppt, pptx, docx, hwp, bat, ps1

Fig. 1. Types of content and files

IoT malware incorporates tactics from current malware, such as polymorphism and anti-analysis, and its operations continue to be those of conventional malware.[9]

However, given the quantity of new IoT devices, it's possible that IoT malware development hasn't yet reached its full potential.

979-8-3503-9747-5/22/$31.00 ©2022 IEEE

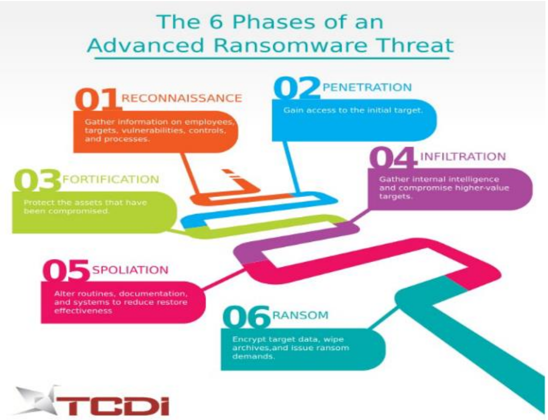

The 6 Phases of an
Advanced Ransomware Threat
01 RECONNAISSANCE
Gather information on employees, targets, vulnerabilities, controls, and processes.
02 PENETRATION
Gain access to the initial target.
03 FORTIFICATION
Protect the assets that have been compromised.
04 INFILTRATION
Gather internal intelligence and compromise higher-value targets.
05 SPOLIATION
Alter routines, documentation, and systems to reduce restore effectiveness
06 RANSOM
Encrypt target data, wipe archives,and issue ransom demands.
TCDi

Fig. 2 phases of advanced ransomware threat

Steps involved in in a ransomware technique, in figure [2]

- Reconnaissance phase
- Penetration phase
- Fortification phase
- Infiltration phase
- Spoliation phase
- Ransom phase

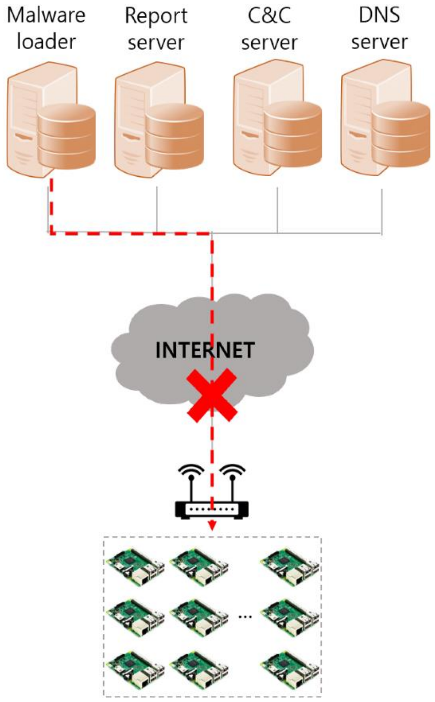


Fig. 3. Malware attack structure

B. DNN Architecture

The Deep neural network architecture is a multi-layered variation of the traditional Ann design. [2] The backpropagation technique underlies both dnn training's two ways; in the forward method, the nonlinear activation function makes calculations from the input to the output layer by layer. In contrast, the derivatives of the error function are computed in the other manner, i.e., in reverse order, from output layer to input layer, in the backward technique. The figure [4] depicts it,

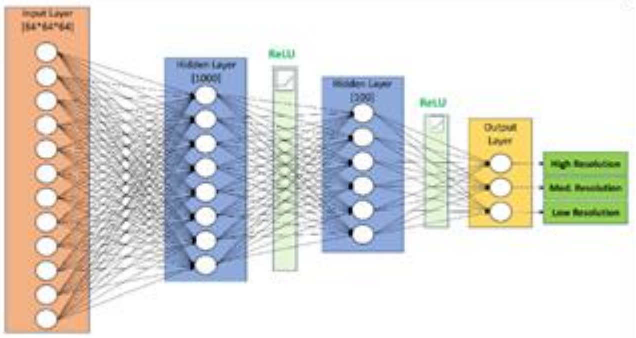


Fig. 4. The DNN Architecture

C. Malware Distribution Network

[3] MDNs are intelligent malicious networks that attackers create in order to carry out ongoing assaults against unidentified people or predetermined organizations.[3] MDNs are several sorts of malicious URLs that include malware that are mutually reliant, and they are constantly in an attack posture dictated by the attacker's approach.

D. Malware Classification framework based on deep learning Algorithm

The use of sophisticated combination and re-construction techniques has allowed malware strains to continually evolve and advance in recent years. Malware identification and categorization has become a big issue for enterprises as malware variations evolve.

According to current scientific and corporate estimates, cyber-attacks cost the world economy trillions of dollars.[4] Malware serves as a tool for attackers to gain access to and abuse the system. Malware is software that conducts unplanned and unwanted/unauthorized operations on the client's behalf.

Viruses, rootkits, trojans, and ransomware are different types of malwares.[4] The main goal of the attack is to influence the victim's computer and steal their private information. Malware detection is a procedure that handles the analysis and identification of dangerous code in software in order to prevent it from running in the client's computer, hence protecting the client's confidentiality, and protecting it from potential viruses. Figure[5] gives us the malware classification.

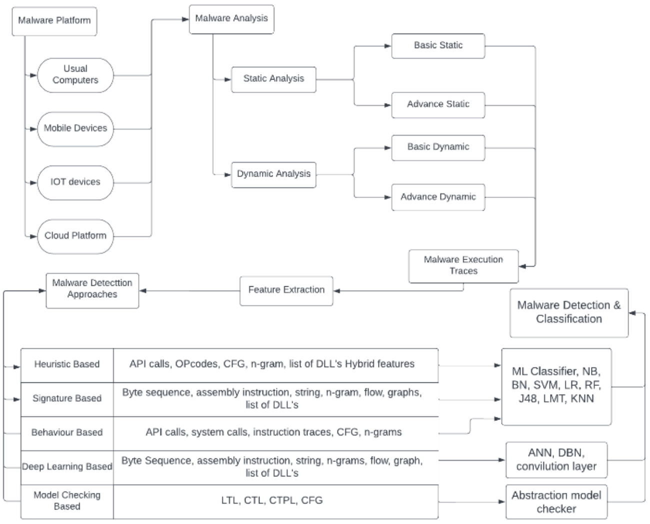

Malware Platform
Usual Computers
Mobile Devices
IOT devices
Cloud Platform
Malware Analysis
Static Analysis
Dynamic Analysis
Basic Static
Advance Static
Basic Dynamic
Advance Dynamic
Malware Execution Traces
Feature Extraction
Malware Detecttion Approaches
Malware Detection & Classification
Heuristic Based
API calls, OPcodes, CFG, n-gram, list of DLL's Hybrid features
Signature Based
Byte sequence, assembly instruction, string, n-gram, flow, graphs, list of DLL's
Behaviour Based
API calls, system calls, instruction traces, CFG, n-grams
Deep Learning Based
Byte Sequence, assembly instruction, string, n-grams, flow, graph, list of DLL's
Model Checking Based
LTL, CTL, CTPL, CFG
ML Classifier, NB, BN, SVM, LR, RF, J48, LMT, KNN
ANN, DBN, convilution layer
Abstraction model checker

Fig. 5. The Malware Classification

E. Types of Malwares

The most prevalent and dangerous malware types that have been spotted recently around us are defined in the below graphic figure [6].

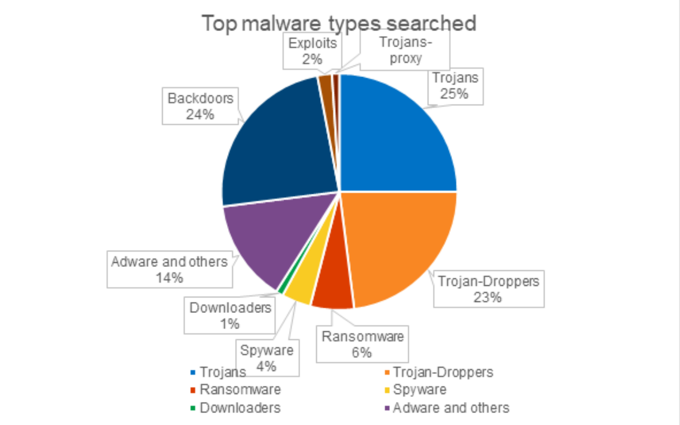


Fig. 6. The Malware Classification

Below discussed a variety of malwares that are seen in & around us.[5]

Adware

Adware, often known as "advertising-supported software," is any program or application that shows adverts, typically through pop-up or pop-under windows. Adware is not necessarily dangerous and is in fact referred to as grayware, despite the fact that they may be annoying to some consumers. As other applications are permitted to operate in the background, adware may cause networks and devices to slow down.

*Browser Hijacke*r

A browser hijacker is a piece of software that adjusts web browser settings without the user's knowledge and redirects the user to websites that the user did not plan to visit. Because it leads the browser to other, frequently harmful websites, it is frequently referred to as a browser redirect virus.

Clickjacking

Clickjacking is a type of assault that tricks people into believing they are clicking on one thing while they are actually clicking on something else. User interface (UI) redressing, which is its other term, more accurately captures the situation. The concealed UI does a separate operation when users click something they believe to be secure.

IP Spoofing

The act of creating Internet Protocol (IP) packets with a fictitious source IP address to masquerade as another computer system is known as IP spoofing, also known as IP address spoofing. IP spoofing enables fraudsters to conduct nefarious activities, frequently undetected.

Phishing

Phishing is the practice of attackers sending malicious emails meant to lead recipients to fall for a scam. Typically, the goal is to persuade individuals to divulge sensitive information such as system login passwords or financial information. Phishing is an illustration of social engineering, a set of strategies con artists employ to influence people's psychology.

Ransomware

Ransomware is a type of virus used to prevent a person or business from accessing files on a computer. Cyber attackers put businesses in a situation where paying the ransom is the quickest and least expensive option to recover access to their files by encrypting these files and requesting a ransom payment for the decryption key.

Trojans

The term "trojan" or "trojan horse" refers to a computer virus. It is a sort of computer program that conceals itself as common applications like utilities, games, and occasionally even antivirus software.

Virus

Malicious software, sometimes known as malware, travels between computers and harms data and software. Computer viruses are one sort of malware. In addition to causing significant operational problems, data loss, and leakage, computer viruses try to destabilize systems. Computer viruses are made to propagate between applications and operating systems, which is an important fact to be aware of.

F. Real time & Adaptive learning malware detection method based on API – pair graph

Malware is increasingly a widespread danger to cybersecurity. [6] Malware is one of the biggest contributors to the destruction of cyberspace since it may be installed on a computer without the owner's knowledge, destroy the machine, and steal information.

Many malware defense solutions relied on two strategies.[6]

1. Static detection, which incorporates sound-assisted heuristic identification and signature-based detection, may quickly identify known malware.
2. Dynamic detection techniques include behavioral detection and pattern matching based detection has a defined range on both known and undiscovered harmful applications. In order to identify malware, deep learning-based detection has now developed into a technique that will be extensively used in the future.

The fundamental sequential algorithm is a clustering technique that is based on a sequence or series of input data that are widely used in the clustering and classification of data.

G. A survey on the cyber security of small-to-medium businesses

A little more than 90% of the world's economy is driven by SMBs. In India, SMBs account for 95 percent of all businesses, making it both a direct and indirect employer. [F] Because small and medium-sized businesses play a significant part in the economy, they are obliged to maintain effective security plans. Attackers are now focusing on SMBs since they are less aware of their vulnerable networks and information resources.[10]

The definition of cybersecurity is "the art of safeguarding the user’s system against hostile attackers, preventing unauthorized script execution, and guaranteeing the confidentiality, integrity, and availability of the information."

The findings of the University study are shown below, with the sorts of assaults most prevalent in SMB and local user systems highlighted.

The National Institute of Standards and Technology (NIST) established the Cyber Security Framework (CSF) to assist enterprises throughout the world in upgrading their critical infrastructure. As per figure [7], The framework includes policies to help organizations avoid, identify, and respond to cyber-attacks.

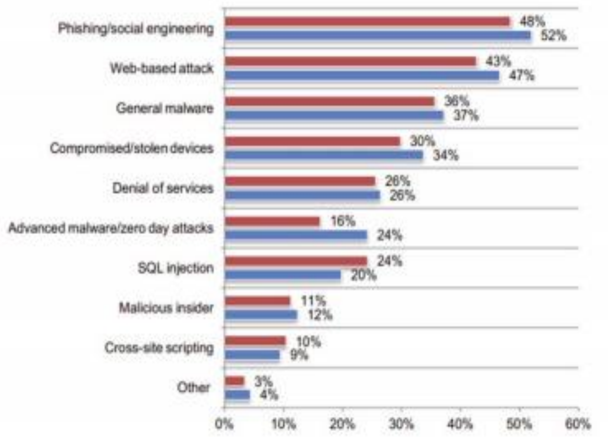


Fig. 7. Types of attack experienced by SMB’s

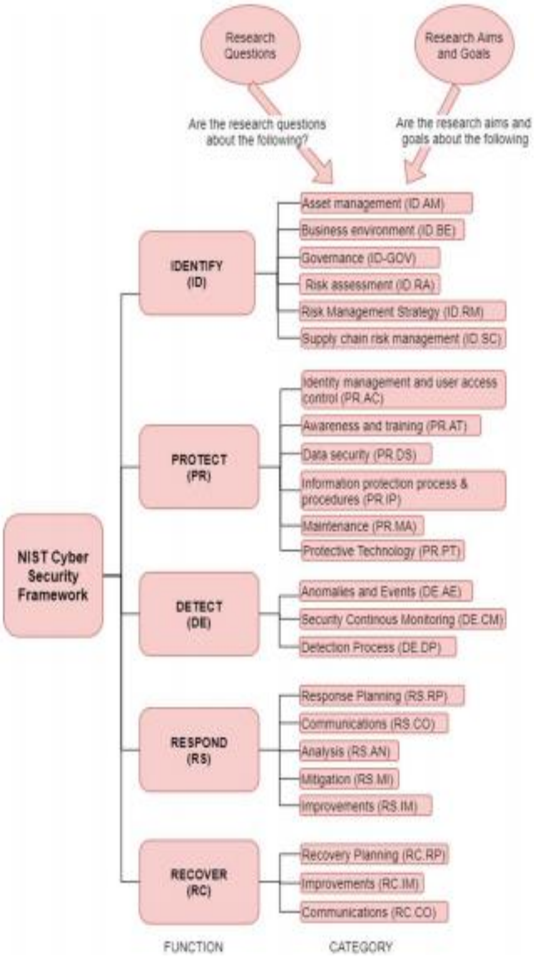

Research Questions
Research Aims and Goals
Are the research questions about the following?
Are the research aims and goals about the following
NIST Cyber Security Framework
IDENTIFY (ID)
Asset management (ID AM)
Business environment (ID BE)
Governance (ID-GOV)
Risk assessment (ID RA)
Risk Management Strategy (ID.RM)
Supply chain risk management (ID SC)
PROTECT (PR)
Identity management and user access control (PR.AC)
Awareness and training (PR.AT)
Data security (PR.DS)
Information protection process & procedures (PR.IP)
Maintenance (PR.MA)
Protective Technology (PR.PT)
DETECT (DE)
Anomalies and Events (DE.AE)
Security Continous Monitoring (DE.CM)
Detection Process (DE.DP)
RESPOND (RS)
Response Planning (RS.RP)
Communications (RS.CO)
Analysis (RS.AN)
Mitigation (RS.MI)
Improvements (RS.IM)
RECOVER (RC)
Recovery Planning (RC.RP)
Improvements (RC.IM)
Communications (RC.CO)
FUNCTION
CATEGORY

Fig. 8. NIST CSF research classification tool

H. DAEMON: Multi-Stage Feature Mining for Dataset/Platform-Agnostic Malware Classification

The objective of malware categorization is to determine which family a harmful version belongs to. Variants from the same malware family exhibit comparable behavioral tendencies.[8] As a result, categorizing recently discovered dangerous programmers and apps aids in determining the hazards they pose. DAEMON is a revolutionary dataset-independent malware classifier. The attributes that DAEMON uses and the way they are mined help to comprehend the majority of the distinctive behavior of the malware family, which helps to make categorization judgments more understandable. Optimized DAEMON employs a big collection of x86 binaries from several malware families aimed targeting Windows-based computers. DAEMON's classification results are extremely accurate across all datasets, demonstrating both its dataset and platform independence. DAEMON can precisely distinguish distinct versions for families whose payloads are also extremely similar.

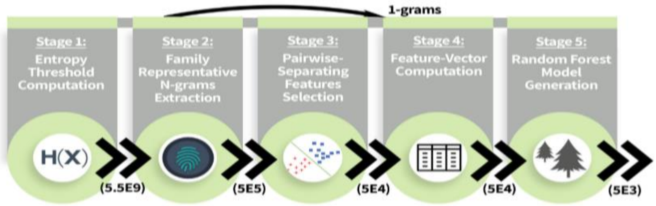

1-grams
Stage 1: Entropy Threshold Computation
Stage 2: Family Representative N-grams Extraction
Stage 3: Pairwise-Separating Features Selection
Stage 4: Feature-Vector Computation
Stage 5: Random Forest Model Generation
H(X)
(5.5E9)
(5E5)
(5E4)
(5E4)
(5E3)

Fig.9. DAEMON’s model generation model

I. RESULTS AND PREVENTION METHODS

A. *Cyber threat hunting techniques*

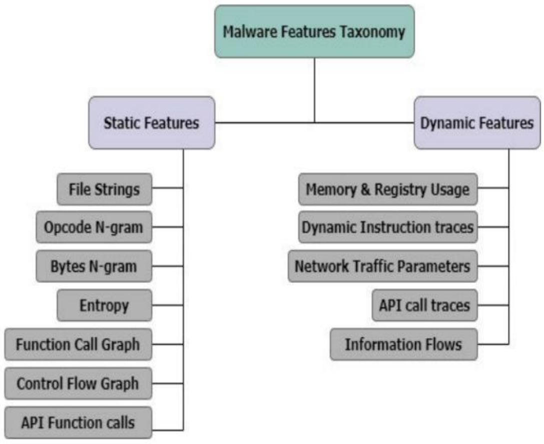


Fig. 10. Malware feature taxonomy

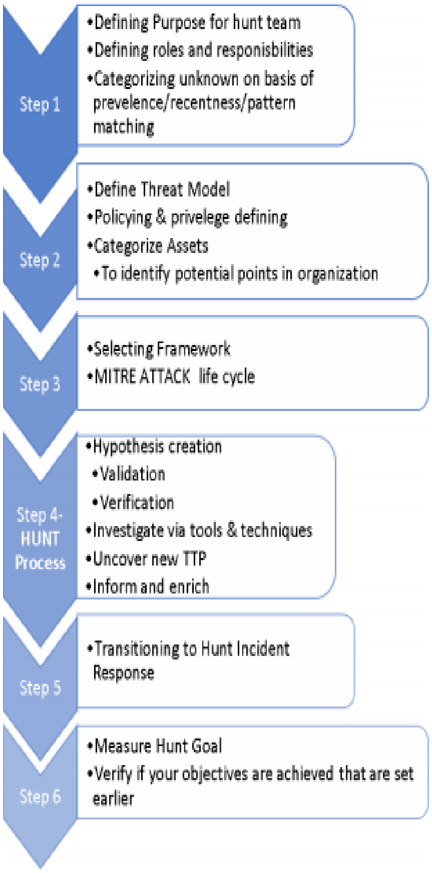

Step 1
•Defining Purpose for hunt team
•Defining roles and responisbilities
•Categorizing unknown on basis of prevelence/recentness/pattern matching
Step 2
•Define Threat Model
•Policying & privelege defining
•Categorize Assets
•To identify potential points in organization
Step 3
•Selecting Framework
•MITRE ATTACK life cycle
Step 4- HUNT Process
•Hypothesis creation
•Validation
•Verification
•Investigate via tools & techniques
•Uncover new TTP
•Inform and enrich
Step 5
•Transitioning to Hunt Incident Response
Step 6
•Measure Hunt Goal
•Verify if your objectives are achieved that are set earlier

Fig. 11. Threat hunting process from scratch

B. *Last Line Defense*

SCADA systems oversee managing critical industrial operations. It is a network comprising several components. Many businesses and SCADA systems rely on vulnerability assessment methods and processes; nevertheless, they are not always appropriate. The newly evolved and more complicated malware, which may be polymorphous, may circumvent protection and acquire access to all necessary information about the industrial system and its activities.

C. *Threat Hunting Framework*

Existing security techniques include firewalls, anti- malware, and Security Information and Event Management (SIEM). These reactive security systems aid in the prevention of numerous forthcoming assaults and also constitute a hesitant approach for recurrent attempts.

The Diamond model of intrusion analysis, the cyber death chain, and the MITRE ATT&CK Matrix are three well-known and commonly utilized threat hunting models.

The MITRE ATT&CK model is a portal that offers threat hunters with attacker tips, tactics, and methods. It also assists the network security team in determining their organization’s security elements and creating and modifying security plans accordingly.

Threat hunting analysis is extensively utilized in the development of Cyber Threat Intelligence (CTI). The data in CTI is used to forecast and classify malware and its activities in the future. Additionally, it gives us access to several fresh classes of vulnerability, keeping users and clients informed.

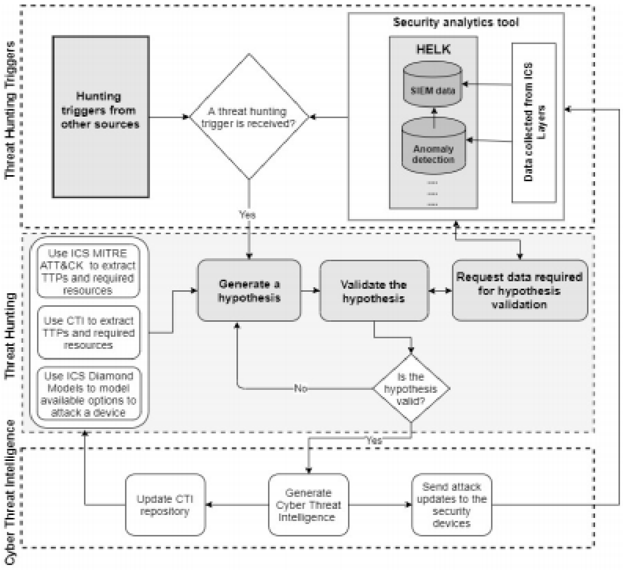
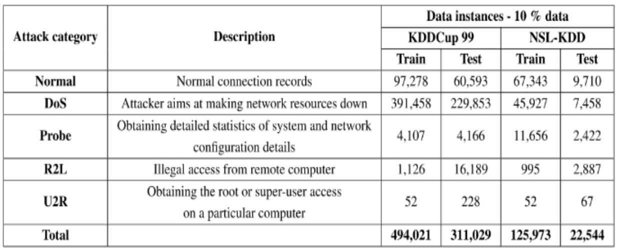

Threat Hunting Triggers
Threat Hunting
Cyber Threat Intelligence
Hunting triggers from other sources
A threat hunting trigger is received?
Yes
Security analytics tool
HELK
SIEM data
Anomaly detection
Data collected from ICS Layers
Use ICS MITRE ATT&CK to extract TTPs and required resources
Use CTI to extract TTPs and required resources
Use ICS Diamond Models to model available options to attack a device
Generate a hypothesis
Validate the hypothesis
Request data required for hypothesis validation
Is the hypothesis valid?
No
Yes
Update CTI repository
Generate Cyber Threat Intelligence
Send attack updates to the security devices

Fig. 12. Proposed ICS threat hunting framework

D. Statistical Analysis



Attack category	Description	Data instances - 10 % data			
		KDDCup 99		NSL-KDD	
		Train	Test	Train	Test
Normal	Normal connection records	97,278	60,593	67,343	9,710
DoS	Attacker aims at making network resources down	391,458	229,853	45,927	7,458
Probe	Obtaining detailed statistics of system and network configuration details	4,107	4,166	11,656	2,422
R2L	Illegal access from remote computer	1,126	16,189	995	2,887
U2R	Obtaining the root or super-user access on a particular computer	52	228	52	67
Total		**494,021**	**311,029**	**125,973**	**22,544**

Fig. 13 Statistical Analysis

E. Robust Intelligent Malware Detection Using Deep Learning

In order to analyse malware, the framework uses a two-stage procedure and applies deep learning to samples of malware that are gathered from end-user hosts. Malware is classified using the first stage mix of static and dynamic analysis. The malwares is then classified into distinct forms of malware in the second step using an image processing method. Numerous experimental evaluations are carried out using different models in benchmark datasets that are made accessible to the public as well as in datasets that are compiled privately. The results demonstrate that deep learning approaches outperformed numerous traditional MLAs. By adding two extra layers to the present architecture, developed frameworks will be able to scale up to analyses a broader variety of malwares in real time.

In a hostile environment, deep learning architectures are susceptible. During the testing or deployment stages, generative network approaches create samples that can readily deceive deep learning structures.

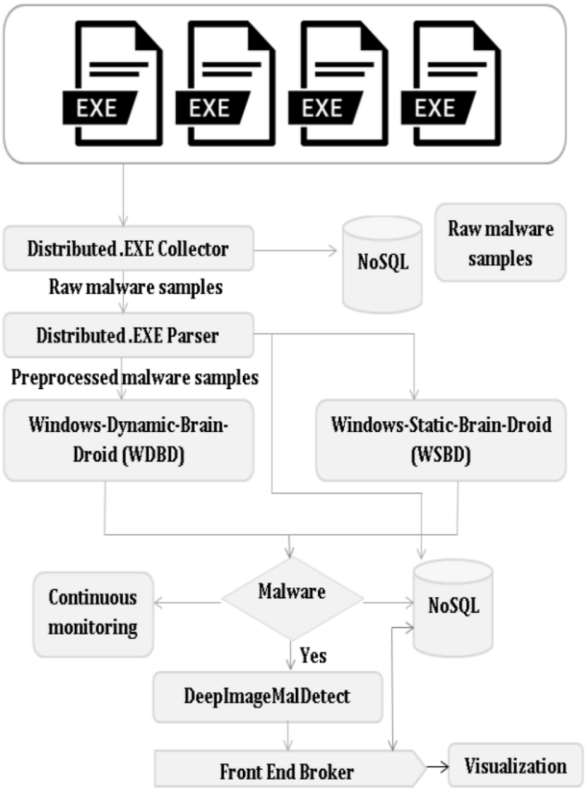


Fig. 14. Proposed deep learning architecture for real-time malware analysis

F. Deep-Ensemble and Multifaceted Behavioral Malware Variant Detection Model

The suggested approach essentially integrates the various criteria to detect malware. Each behavioral characteristic can identify a specific aspect of the file's goodness or malice.

Automatic extraction of a deep hidden features vector from a deep sequential model's final hidden layer. This hidden representative feature is retrieved from the hidden layer of each learned deep learning model and then concatenated into a single feature vector that is fed into the XGBoost algorithm for training a collection of ensemble classifiers.

While lowering the false positive rate is crucial, it is not as important for malware detection as lowering the false negative rate. Reducing the FNR is a crucial security criterion in malware detection since it may lead to successful assaults on the system.

The suggested model MB-MVDS-XGB achieves the best decrease in terms of FNR, followed by the feature vector paired with the sequential deep learning MB-MVDS-SDL, which achieves a 0.67% negative rate, as shown in the picture below.

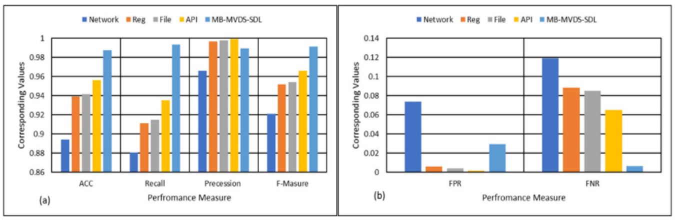


Fig. 15. The performance comparison with the related work

IV. CONCLUSION

On further investigation of the preventive techniques, the antivirus should be able to analyze the resource in seconds and block the user before they are impacted, and it should be synchronized with the updates for greater safety. The antivirus is a cloud-based program that may be used from any location. This antivirus will provide us with a greater level of security.

Our work illustrates the various malware and viruses in the statistics and with graph support by assessing current malware along with its distribution network, using DNN architecture, DAEMON, deep learning, threat hunting frameworks, API pair graph and by analyzing surveys of various companies, digital healthcare. Malware is a harmful program produced for illicit activities and has now become the themes the studies are working on. Both static and dynamic analysis are used while analyzing numerous categories. The comprehensive investigation was carried out in order to decrease false positives and better identify the core cause of the virus.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Meenakshi Sundaram B. for guiding us throughout this paper and helping us to come up with a sophisticated model by pushing us to what we can achieve. We are glad we got to work on this paper and make it come to life.

REFERENCES

[1] “Comprehensive Analysis of Today’s Malware and Its Distribution Network: Common Adversary Strategies and Implications “SIWON HUH, SEONGHWAN, JINHO CHOI, SEUNGWON SHIN AND HOJOON LEE

[2] "Robust Intelligent Malware Detection Using Deep Learning", R. VINAYAKUMAR, MAMOUN ALAZAB, K. P. SOMAN, PRABAHARAN POORNACHANDRAN, AND SITALAKSHMI VENKATRAMAN

[3] "Potential Risk Analysis Method for Malware Distribution Networks" DOHOON KIM

[4] "A New Malware Classification Framework Based on Deep Learning Algorithms",ÖMER ASLAN , ABDULLAH ASIM YILMAZA.

[5] Computer Viruses and Malware by John aycock

[6] "A Real-Time and Adaptive-Learning Malware Detection Method Based on API-Pair Graph", SHAOJIE YANG , SHANXI LI, WENBO CHEN, AND YUHONG LIU

[7] "A Survey on the Cyber Security of Small-to-Medium Businesses: Challenges,Research Focus and Recommendations"
ALLADEAN CHIDUKWANI , SEBASTIAN ZANDER , AND POLYCHRONIS KOUTSAKIS

[8] "DAEMON: Dataset/Platform-Agnostic Explainable Malware Classification Using Multi-Stage Feature Mining",RON KORINE AND DANNY HENDLER.

[9] S. K. B V, S. Sharma, K. S. Swathi, K. R. Yamini, C. P. Kiran and K. Chandrika, "Review on IoT based Healthcare systems," 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), 2022, pp. 1-5, doi: 10.1109/ICACTA54488.2022.9753547.

[10] M. Gupta, J. Dhanush, R. Vikas, S. B. V. Krishna, A. R. Naik and C. Gowda, "A Safe and Reliable System for Monitoring the Home Remotely," 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 1-4, doi: 10.1109/ICCCI54379.2022.9740879.