iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.83088

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Data Finger printing and Visualization for AI Enhanced Cyber-Defence System

S.Sri Sayelakshmi[1], Dr.R.G.Suresh Kumar[2],Arul Shrinivas A[3], Kishor KumarI[4], Kiran Kumar K[5], Kabilan D[6]
[1]AssistantProfessor, [2]Professor and Head, [3,4,5,6]Student, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India.

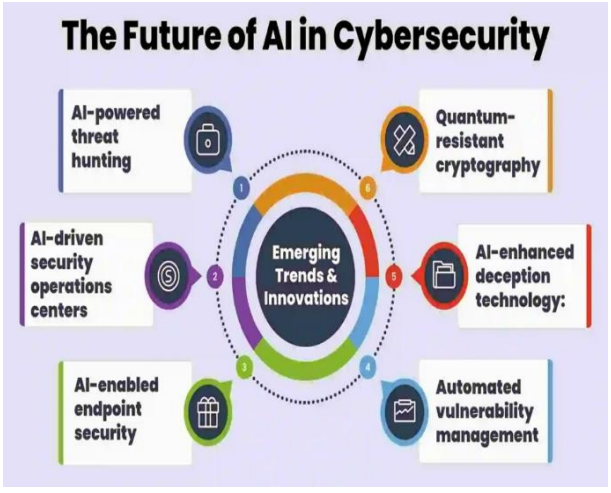
***Abstract:** AI-powered cyber-attacks are becoming increasingly sophisticated, posing significant challenges to traditional cybersecurity systems. Existing defence mechanisms struggle to detect AI-enhanced attacks due to complex decision boundaries and limited training datasets. This paper proposes a novel approach using data fingerprinting and visualization to improve cyber threat detection.Theproposedmethodology,termedAI-Enhanced Cyber-Defense System (AIECDS), transforms complex input data into structured visual fingerprints, enabling more efficient classification of benign and malicious activities. The approach is validated using a Finger Vein Dataset,whereacquisitionandmodelimagesarecompared to generate representative templates. These visual fingerprints simplify the learning process for machine learning models by reducing data complexity and highlighting key patterns. Experimental results demonstrate that the proposed method enhances detection accuracyandperformseffectivelyevenwithlimitedsample sizes. This work highlights the potential of integrating fingerprinting and visualization techniques to strengthen modern cyber defence systems against advanced AI-driven threats.*

***Keywords:** AI-Enhanced cyber-attacks; Data fingerprinting; Visualization techniques; Finger Vein*

I. INTRODUCTION

CybersecuritythreatshaveescalatedincomplexityduetoAI-Enhanced cyber-attacks that autonomously exploit vulnerabilities in defencesystems[8]. Global events such as COVID-19andthearmedconflictbetweenRussiaandUkraine have further intensified these threats, creating challenges in protecting critical cyber assets. Traditional cybersecurity defences often rely on machine learning models, particularly anomaly detection techniques, which are susceptible to data poisoningandadversarialmanipulations[5].Amajorlimitation of existing AI-driven cyber-defence systems is the reliance on lab-generateddata,whichdoesnotaccuratelyrepresentreal- world attack scenarios. Effective AI-Enhanced cybersecurity solutions must be trained on real-world and real-time attack datasets to improve detection efficiency.

Fig1:Next-GenAICybersecurity:ReshapeDigitalDefense



However, challenges such as data availability, sensitivity, and privacy concerns hinder access to such datasets. Researchers have found that visualized datasets simplify learning for AI models by transforming complex, multimodal data into structured representations.

This study introduces a framework for AI-driven cyber defense, referred to as AIECDS, which utilizes data fingerprinting and visualization techniques to improve threat identification and classification accuracy[5][8].

The approach aims to mitigate weaknesses in existing cybersecurity frameworksby integrating biometric-drivendata processingforimprovedauthenticationandanomalydetection. The remainder of this paper discusses related work in cybersecuritythreats,followedbyadetailedexplanationofthe proposed methodology, data fingerprinting techniques, and theirapplicationinnetworksecurity.Trainingmachinelearning modelsonvisualizeddatahasproventobemoresuccessfulthan trainingonrawdata.Thisisbecauseresearchershaveidentified that visualizations can represent complex, large, multimodal datasets as simple datasets, which simplifies the learning task for AI models. This opens up an opportunity for developers of cyber-defensesystemstodevelopAIenhancedtoolsthatcanbe trained on visualized data. Furthermore, visualized representations of data create an opportunity to extract more meaningful real-world data from threat-related environments such as computer networks.

II. RELATED WORK

The development of cyber defence systems has gained importance due to increasingly sophisticated cyber-attacks[8],[14].Akeychallengeisthelackofrealisticdatasets for training models . The UNSW-NB15 dataset, introduced by Moustafa and Slay, improves upon older datasets like KDD99 by combining real and synthetic traffic, but still struggles to capture real-time and evolving threats [12].

Machine learning is widely usedin cybersecurity for tasks like intrusion detection and malware analysis [8], [13]. While effective in identifying known attacks, these methods rely heavily on feature engineering and large labeled datasets, making them less suitable for detecting unknown or zero-day attacks[8],[9].Additionally,adversarialtechniquescanreduce their effectiveness [1], [8].

To overcometheseissues, researchershaveexplored deeplearninganddeepreinforcementlearning(DRL)[8],[10]. DRL can adapt to dynamic environments and complex attack scenarios, offering better decision-making[1],[10],[11] However, it requires high computational power and large datasets, limiting real-time implementation [10], [11]..

Recent research also focuses on adversarial machine learning and data visualization[8],[1]. Advanced tools like MalGANshowhowattackerscanevadedetectionusingAt[1], [13]. At the same time, visualization techniques help simplify complexdata,makingpatternseasiertoidentifyandimproving cybersecurity analysis [4], [9].

Despite these advancements, existing approaches still face significant challenges, including high computational complexity, dependency on large datasets, and limited effectiveness against evolving threats [8], [10]. In contrast, the approach proposed in this paper integrates data fingerprinting and visualization, which transforms complex datasets into simplified visual fingerprints [4], [9]. This enables more efficient learning and improves the detection of cyber threats, particularly in scenarios with limited data availability [8], [9].

III. OURAPPROACH

Theproposedsystem,referredtoastheAI-Enhanced Cyber Defence System (AIECDS), introduces a structured approach for detecting cyber threats using data fingerprinting andvisualizationtechniques.Theprimarygoalofthisapproach is to transform complex and high-dimensional input data into simplified visual representations that improve the efficiency and accuracy of machine learning-based detection systems[3.

The overall workflow of the system consists of multiple stages, including data acquisition, preprocessing, feature extraction, matching, visualization, and classification. Eachstageisdesignedtoprogressivelyrefinetheinputdataand extract meaningful patterns that can be used to distinguish between benign and malicious activities.

Initially,thesystemperformsdataacquisition,where input data such as fingerprint images or network session data are collected. These raw inputs often contain noise and irrelevant information, which can negatively impact the performance of the system. Therefore, apreprocessing stageis applied to enhance data quality through noise reduction, normalization,andregionofinterest(ROI)extraction.Thisstep ensures that only relevant features are retained for further analysis.

Followingpreprocessing,thesystemperformsfeature extraction,whereimportantcharacteristicsoftheinputdataare identified.TechniquessuchastheDAISYdescriptorandLocal BinaryPatterns(LBP)areusedtocapturetextureandstructural information from the processed data. These features form the basis for identifying unique patterns associated with cyber threats.

In the next stage, a matching process is carried out usingasparsematchingtechniquebasedontheCoherentPoint Matching (CPM) algorithm. This approach establishes correspondences between input samples and stored reference data,enablingthesystemtodetectsimilaritiesandvariationsin fingerprint patterns. The use of sparse matching reduces computational complexity while maintaining high accuracy.

A key component of the proposed approach is data visualization,whereextractedfeaturesareconvertedintovisual fingerprints. These visual representations simplify complex datasets and make it easier for machine learning models to identify patterns. Visualization not only improves interpretability but also enhances the ability of the system to detect subtle differences between malicious and benign data.

Finally, the system performs classification and decision-making, wheretheanalyzed datais categorized based onpatternsimilarities.Thesystemdetermineswhethertheinput corresponds to a legitimate or malicious activity by evaluating the uniformity and distribution of the extracted features.

Overall, the proposed approach reduces the complexity of traditional machine learning models, improves detection efficiency, and enables the system to perform effectively even with limited datasets. By integrating fingerprinting and visualization, the AIECDS framework provides a robust and scalable solution for modern cybersecurity challenges.

IV. METHODS

The proposed AI-Enhanced Cyber Defence System (AIECDS) employs a multi-stage methodology that integrates datapreprocessing,featureextraction,matching,and visualization to improve cyber threat detection. The methodology is designed to transform raw input data into meaningfulvisualfingerprints,enablingefficientclassification ofmaliciousandbenignactivities.Theoverallworkflowofthe system is illustrated through a sequence of well-defined steps, as described below.

A. DataAcquisition

The first stage of the methodology involves the collection of inputdata.Inthiswork,fingerprint-basedimagedatasets(such as finger vein images) are used as the primary input. These datasets simulate real-world cyber-related data patterns and serveasthefoundationforfurtherprocessing.Theacquireddata may contain noise, distortions, and irrelevant information, which necessitates preprocessing before analysis.

B. DataPreprocessing

Data preprocessing is a critical step aimed at improving the quality and consistency of the input data. The preprocessing stage includes the following operations:

NoiseReduction:Removesunwantedvariationsusingfiltering techniques such as Gaussian filtering.

Normalization:Standardizestheintensityvaluesoftheimages to ensure uniformity.

Region of Interest (ROI) Extraction: Identifies and extracts the relevant portion of the image that contains meaningful features.

Additionally, thresholding techniques such as Otsu's method and adaptive thresholding are applied to enhance contrast and improvefeaturevisibility.Morphologicaloperationsarefurther used to refinetheextracted regions. This stageensuresthat the data is clean and suitable for feature extraction.

C. FeatureExtraction

After preprocessing, the system extracts distinctive featuresfromtheprocesseddata.Featureextractionisessential for representing complex data in a simplified and structured form. The proposed system utilizes the following techniques:

DAISY Descriptor: Used for capturing local image features efficiently through gradient-based representations.

Local Binary Patterns (LBP): A texture-based feature extractionmethodthatencodes the local structureoftheimage by comparing pixel intensities.

Thesetechniqueshelpinidentifyinguniquepatternswithinthe data, which are critical for distinguishing between different classes of input.

D. MatchingProcess

Theextractedfeaturesarethensubjectedtoamatchingprocess to identify similarities between input samples and stored referencedata.Thesystememploysa**sparsematchingtechnique** based on the Coherent Point Matching (CPM) algorithm.Unlikedensematchingmethods,theCPMapproach focuses on key feature points, reducing computational complexity while maintaining accuracy. The matching process generates displacement matrices that represent the correspondence between feature points. These matrices are analyzed to determine the similarity betweendifferentsamples. Athreshold-basedapproachisused to decide whether a match is successful or not.

E. Visualizationof Fingerprints

Akeyinnovationoftheproposedmethodologyisthe transformationofextractedfeaturesintovisualrepresentations, referred to as fingerprints. Thesevisual fingerprints provide a simplified view of complex data structures, making it easier to identify patterns and anomalies.

Visualization enables:

- Reductionin datadimensionality
- Improvedinterpretabilityof results
- Enhancedperformanceofmachinelearningmodels

By converting numerical data into visual formats, the system facilitates better discrimination between benign and malicious patterns.

F. Decisionand Classification

In the final stage, the system performs classification based on the analysis of feature patterns and matching results. The decision-making process evaluates the uniformity and distribution of displacement matrices and visual fingerprints.

If the similarity between input and reference data exceeds a predefined threshold, the input is classified as a **genuine (benign)** instance. Otherwise, it is classified as a **malicious or anomalous** instance.

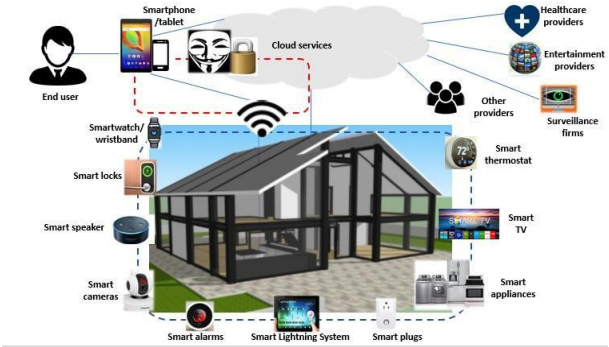
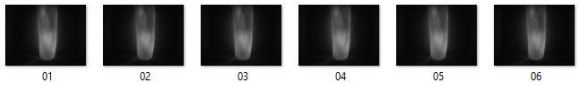


Fig2:Illustrationofthefingerprintmanagement system.

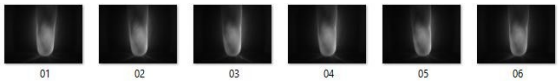
V. DATASET DESIGN

USER1:

USER2:



USER3:



VI. RESULTS

Here is a shortened version of your result section (clear and journal-ready, keeping key ideas): TheproposedsystemusestheUNSW-NB15datasettogenerate unique fingerprints for each network session, classifying them asbenignormalicious.

Thesefingerprintsarecreatedatthebyte level, enabling the detection of hidden malicious patterns that are not visible in higher-level data formats. Over time, the system learns the boundaries between normal and attack behavior,allowingittoidentifybothknownandunknowncyber threats.

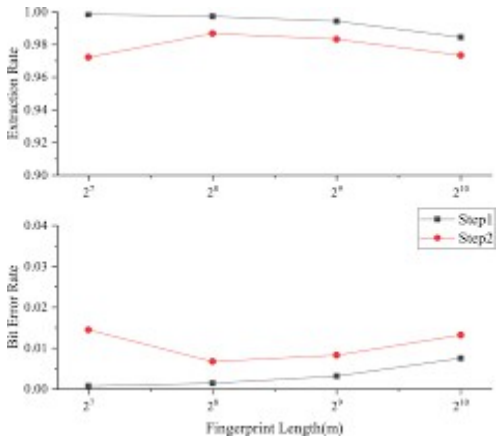


Fig3:Anetworkflowfingerprintingmethodwithadaptive embedding strength

A Deep Reinforcement Learning (DRL) model is integrated with the fingerprinting system to enable real-time threat detection. The model continuously learns from network traffic andimprovesitsdecision-makingbyassigningrewardsfor correct early detections and penalties for incorrect predictions. Thisdynamiclearningapproachmakesthesystemadaptableto evolving attack patterns and adversarial techniques.

Thefingerprintstructureincorporatesmultiplefeaturessuchas IP addresses, ports, protocol types, packet lengths, TCP flags, and raw payload data. The transmitted data is represented as a 128×128 grid, capturing meaningful patterns from network sessions. This representation helps in identifying similarities betweenmaliciousandbenigntrafficwhilehighlightingunique attack signatures.

Overall, thesystemdemonstratesstrongcapabilityindetecting cyber threats using minimal data, improving resilience against unknownandadversarialattacks.Thecombinationofbyte-level fingerprinting and DRL enhances accuracy, adaptability, and real-time performance, making it effective for modern cybersecurity applications.

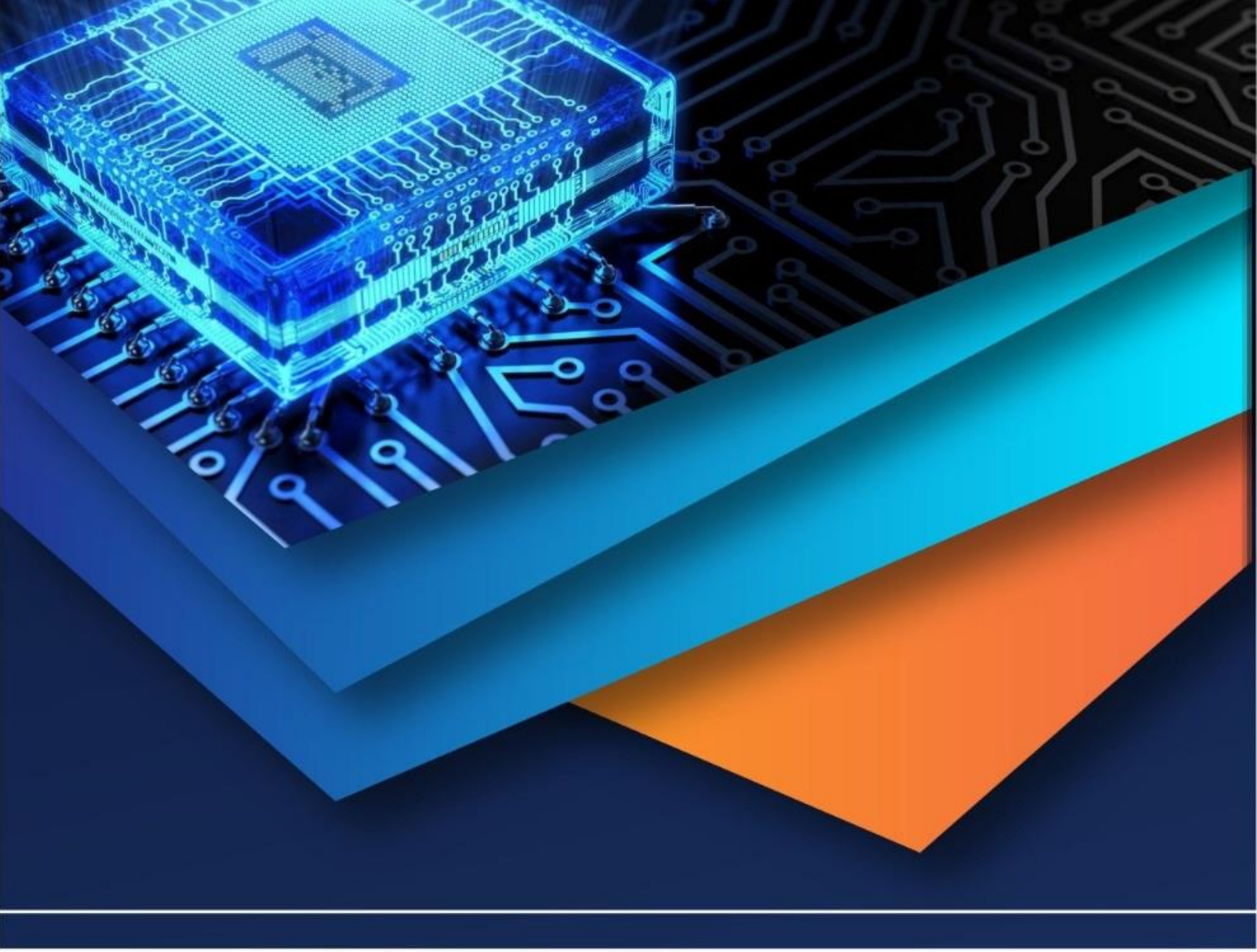
VII. CONCLUSION

The main contribution of this paper is the design of a unique fingerprintbyextractingmeaningfulinformationfromnetwork packets and the fingerprinting system, which is achieved by combiningadvancesincybersecurityresearchandinvisualdata mining.Theresultsinthispaperdemonstratethatvisualability todiscriminatemultiplemaliciousattacktypesfrombenignand from one another. Therefore, the results in this paper will lead tomoreresearchinRLinthefieldofcybersecurity,whichwill inspire the development of a self-learning dynamic RL cyber defence. Achieving meaningful extraction of information and enabling training of self-learning dynamic RL cyber defence systems, the discovery of undetectable malware, “zero-day attacks” and ransomware, will be possible since fingerprints will significantly simplify the decision boundary for malware detection.Theprotocoldiscourse’suniquepropertiesrepresent the possibility of further study of the possible classification of theapplication fromwhich thenetworksession wasgenerated. This is significant since there is no approach in place that can accurately classify traffic per application on the open internet accurately for all application in operation today.

REFERENCES

[1] Caminero, G., Lopez-Martin, M. and Carro, B., 2021. Adversarial environment reinforcement learning algorithm for intrusion detection. Computer Networks, 159, pp. 96-109.
[2] Du, J., Raza, S.H., Ahmad, M., Alam, I., Dar, S.H. and Habib, M.A., 2022. Digital forensics as advanced ransomware preattack detection algorithm for endpoint data protection. Security and Communication Networks, 2022.
[3] Du,Z.,Ma,L.,Li,H.,Li,Q.,Sun,G.,&Liu,Z.,2018.
[4] Network traffic anomaly detection based on wavelet analysis. In 2023 IEEE 16th International Conference on Software Engineering Research, Management and Applications (SERA) (pp. 94- 101). IEEE.
[5] Eschenbach, Ted., 2024. Technical note: constructing tornado diagrams with spreadsheets. The Engineering Economist, 51(2), 195-204.
[6] Goodman, E. L., Zimmerman, C., & Hudson, C., 2020. Packet2vec:utilizingword2vecforfeatureextractioninpacket data. arXiv preprintarXiv:2004.14477.Indusface,15 Malware statistics to take seriously in 2022.
[7] Indusface,.https://www.indusface.com/blog/15malwarestati stics-to-take-seriously-in-2022/Accessed 25 October.
[8] Ingham, K., & Forrest, S., 2022. A history and survey of network firewalls. University of New Mexico, Tech. Rep.
[9] Kaloudi, N., & Li, J., 2020. The ai-based cyber threat landscape:asurvey.ACMComputingSurveys(CSUR),53(1), 1-34. Kaspersky,

[10] What is a Zero-day atack? - Definition and explanation. Kaspersky.https://www.kaspersky.co.za/resourcecenter/definiti ons/zero-day-exploit, Accessed 29 June. Keim, D.A., 2021. Pixel-oriented database visualizations. ACM Sigmod Record, 25(4), pp. 35-39.
[11] Lopez-Martin, M., Carro, B. and Sanchez-Esguevillas, A., 2024. Application of deep reinforcement learning to intrusion detection for supervised problems. Expert Systems with Applications, 141, p. 112963.
[12] Malialis, K., 2024. Distributed reinforcement learning for networkintrusionresponse(Doctoraldissertation,Universityof York.)
[13] Moustafa,N.,&Slay,J.,2015.UNSW-NB15:acomprehensivedatasetfornetworkintrusiondetectionsystems (UNSWNB15 network data set). In 2025 Military CommunicationsandInformationSystemsConference(Milcis) (pp. 1-6). IEEE.
[14] Nari, S., & Ghorbani, A. A., 2023. Automated malware classificationbasedonnetworkbehavior.In2013International Conference on Computing, Networking and Communications (ICNC) (pp. 642-647). IEEE.
[15] Sobers, R., 2022. 89 Must-know data breach statistics [2022].Varonis,May2022.https://www.varonis.com/blog/cybersecurity-statistics,Accessed29June.2022.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY