IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

REVIEW AND ANALYSIS ON IDENTIFY MALICIOUS ACTIVITY ON A SECURITY DATA BREACHES BASED ON A COMPUTER SYSTEM

[1]BANAVATHU GUNA RATNA JASWANTH NAYAK, [2]SOUJANYA BHUKYA

Regd No: 206301031,IVth B.tech (Computer science and engineering) , Faculty of engineering ,

Gurukul kangri Vishwavidyalaya Haridwar.

Lecturer in Computer Science,SVRK GDC (M) , Nidadavole,E.G.District ,A.P

ABSTRACT

Malware is a major problem for everybody using the internet today. Malware that can quickly change its behaviour in response to new environments or security measures is called polymorphic malware. In order to evade detection by signature-based malware detection methods, polymorphic malware is continually changing its characteristic qualities. We employed several machine learning strategies for detecting malware and other dangerous threats. The high detection ratio is proof that the system employs the most effective algorithm possible. Accuracy information was enhanced by the confusion matrix's measurement of false positives and false negatives. Modern malware is highly developed and intricately designed to cause maximum damage. Most sophisticated malware is resilient and can evade detection for long periods of time. In this research, we'll be looking at a particularly sophisticated form of malware called advanced persistent threats (APTs). It is possible to attribute most cases of cyber espionage and sabotage to APTs. Complex, custom-tailored, and stealthy up until the point of compromise, advanced persistent threats (APTs) are difficult to detect. Automated, target-specific malware is deployed within a host or network by APTs so that assaults can be launched on demand in response to constant monitoring. In this paper, we will summarise and evaluate the state-of-the-art approaches to detection from several fields of study.

1. INTRODUCTION

In the world of modern technology, cyberattacks are currently the most important threat that we face. Harnessing a system's flaws for malicious ends like theft, modification, or destruction is what this word refers to. Malicious software is an example of a cyberattack. To cause harm to a computer, user, business, or computer system, malware is any programme or collection of instructions [1]. Malicious software (or malware) is another name for this type of programme. The term "malware" refers to a wide variety of malicious software, including as viruses, Trojan horses, ransomware, spyware, adware, rogue software, wipers, scareware, and so on. "Malware" is shorthand for "rogue software." Malicious software is defined as any code that is executed without the user's knowledge or agreement [2].

This study demonstrated that malicious traffic on computer systems can be detected and network security can be improved by combining the results of malware analysis and detection with machine learning algorithms to compute the difference in correlation symmetry integrals (using Naive Byes, SVM, J48, RF, and the proposed approach). Machine learning methods were applied to the findings from malware analysis and detection in order to determine the degree of asymmetry in the correlations between the two.

To assess whether or not a piece of software or network connection poses a security concern, malware detection modules analyse collected and trained data [3,4]. Let's pretend for the sake of argument that machine learning systems can explain the rationale for the patterns they've noticed [5]. Machine learning-trained algorithms can improve their predictive abilities by considering and responding to feedback about how they performed on past projects [6].

Cybercriminals employ dangerous software and steal critical information to threaten individuals, organisations, and governments worldwide [7]. Every day, thousands of dishonest people try to use malicious software to gain unauthorised access to networks, steal sensitive data, or launder money. As a direct result, safeguarding sensitive data has emerged as a pressing issue for the scientific community. Using data mining and machine learning classification techniques, this study aimed to develop a complete framework for the detection of malware and the protection of sensitive data from hackers. In this study, we investigate signature-based and anomaly-based characteristics to develop a reliable and efficient method for malware classification and detection. Experiments have shown that the suggested strategy outperforms existing alternatives [7].

Malware that targets modern websites has become increasingly widespread and sophisticated, posing a significant risk to the online safety of these sites [8]. The digital environment, often known as cyberspace, is subject to a variety of cyberattacks, which are depicted in Figure 1. Malware is computer software that was developed specifically for the aim of wreaking havoc on a computer or network in some way, such as by spying on its users or stealing money from them. Malware is increasingly targeting critical infrastructure including the systems that run the internet of things, medical devices, and the controls for factories and factories' environmental systems. Modern spyware is extremely difficult to detect since it continuously alters its code and methods of operation. As malicious software has become more widespread, traditional defences that rely on signatures have been ineffective. Instead, it is required to implement a wider variety of preventative measures [9].

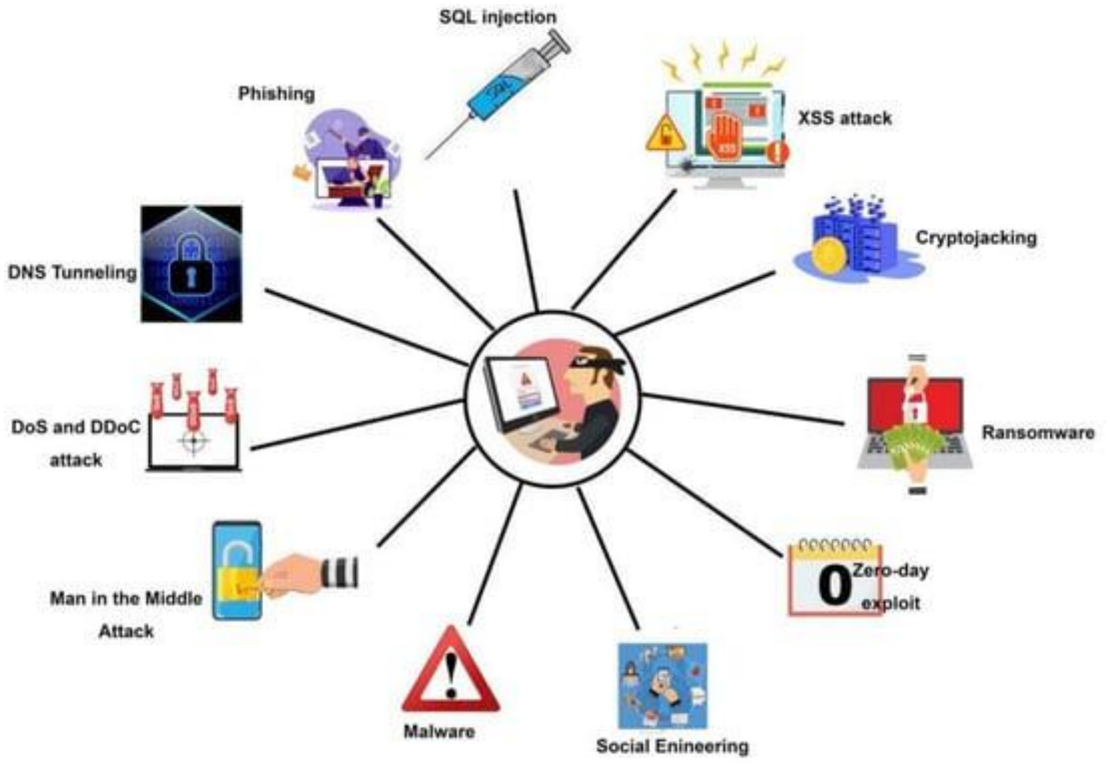

Figure 1. Types of cyberattacks.

Malware families [10] can be characterised by their shared behaviour traits using either static or dynamic learning methods. Dynamic analysis takes into account the behaviour of potentially hazardous files, as opposed to static analysis, which simply assesses their contents without actually running them. To achieve this, monitoring code is added to dynamic binaries, data flows are tracked, and function calls are recorded [11]. Such static and behavioural artefacts may be used by machine learning algorithms to better comprehend the ever-changing nature of today's malware. In this way, the algorithms would be able to identify sophisticated malware attacks that could otherwise elude signature-based detection methods. Because they do not rely on signatures, solutions that are based on machine learning have a higher success rate against freshly released malware. In order to acquire and represent features accurately, it is helpful to use deep learning algorithms that are capable of doing so on their own [12].

2. LITERATURE REVIEW

The community concerned with the detection and classification of malware has published a number of review papers in an effort to determine specific definitions for the many forms of malware, the stages of growth that malware must pass through in order to become more complex, and the obfuscation techniques that malware uses. Static, dynamic, and hybrid methods of analysis were presented by the authors of [13,14]. Signature-based, heuristic-based, specification-based, machine learning-based, deep learning-based, and multimodal-based detection methods are all distinct categories created by the authors. Some of the most difficult problems that researchers in the field of malware detection have encountered include class imbalance, open and public benchmarks, idea drift, adversarial learning, and interpretability of models. In addition, the authors of reference [15] included a sandbox detection strategy to the strategies that were already covered in their assessment of malware detection methods.

Furthermore, in [16], the distinction between static, dynamic, and hybrid feature extraction approaches was developed, along with a relationship between the data types most suited to each method. In addition, both the method of feature extraction and the process of feature selection are broken down and presented from a variety of angles in [17]. The phases of machine learning-based malware detection models are outlined in [18], wherein their authors undertake a review of such models. The authors organised their presentation of the feature extraction process by the sorts of data that were actually extracted.

N-gram, graph, and dataset-based feature extraction approaches are highlighted in the classification of feature extraction methods presented in [19]. Another source that analysed the literature extensively and categorised feature extraction strategies into static, dynamic, and hybrid approaches is reference [20]. This grouping was a direct result of the analysis techniques used. They looked into the analysis process from two perspectives for their review: the phases and the strategies. The steps of analysis were broken down into three categories: representation, assembly, and categorization.

Furthermore, it was concluded that the family analysis, the analysis of similarities, and the study of variants were all valid approaches to analysis. The goal of this research was to show how different selection and detection tactics affect the malware detection model's overall efficiency. As a result, we have investigated the effect of several feature types and classification algorithms to show that no single approach can reliably identify all forms of malware. When it comes to the data-driven feature selection step, the author provided a thorough overview of current approaches and methodologies. Based on the similarities between them, they classified the feature selection procedures as either theoretical, spare learning, or statistical.

Some of the researchers concentrated their efforts on analysing the development of modern sophisticated malware as well as the challenges that have arisen as a result of this evolution. In [21], the authors address the second generation of malware, including their development stages, and analyse the progression of malicious with relevant detection approaches. We classified these approaches as signature-based, behavior-based, heuristic-based, specification-based, energy-based, bio-inspired, and machine/deep learning-based to show how malware evolution and detection methods are intertwined.

However, they only considered API calls in their study, thus their results don't account for additional anti-analytical activities that necessitate an understanding of ot. According to [22], the ratio of evasive malware, the trend of employed evasion strategies, and the impact of anti-analysis operations on the analysis and detection procedures are all discussed after analysing the FFRI dataset to determine these factors.

Malware that may evade detection is inherently more complex. Some researchers were able to avoid being caught by cutting-edge malware detection when it was carried out in a particular manner or with a particular strategy. Multiple facets of malware attacks on home automation systems were looked into by the study's authors [23]. The assaults were classified into categories based on characteristics such as the target's smart home's architecture, its central processing unit, and its level of physical security. The vulnerabilities of VPN filter malware, its effects on router makers, and its possible impact on the smart home network are also examined.

Similarly, the characteristics, models, payload delivery mechanisms, and advanced evasion tactics of advanced persistent threats (APTs) were examined in depth in the study cited as Ref. [24]. Virus-fighting analytical approaches and contemporary application hardening methodologies have also been classified. This was done to make things simpler to organise. It was claimed that the measures intended to create a safe area free of APTs had been put into effect.

However, in Ref. [25], the authors conduct a comprehensive review of data mining-based malware detection methods. Frameworks were provided that describe the use of machine learning in both signature-based and behavior-based approaches to malware detection. In addition, a breakdown of the issues was offered, and data on how often machine learning techniques were used in the research was shown. An overview of a data mining-based malware detection method is available in [26], which you can access here. In addition to the technique that is based on signatures, they included the ways that are based on heuristics and specifications for detecting malware. Furthermore, they discussed the benefits and drawbacks of each of the approaches that were described.

However, this research gives an extensive review that draws from a wide range of published works to create a classification system for malware analysis and detection methods. Furthermore, this study highlights the data types that are most frequently employed for each of these approaches. In addition, in contrast to the existing taxonomies, which only cover general methods of detection such as signature and behavioural, the one presented in this review contains a more in-depth taxonomy that introduces the known methods of detection in greater depth. It achieves so by classifying novel detection methods into novel subcategories and linking those novel subcategories to the most frequent data kinds. This affords academics a chance to hone their expertise of the many detection strategies already in use. This survey presents the feature extraction phase from the perspective of which technique is used to achieve the extraction process, thereby introducing a clearer concept that emphasises the differences between the data collection process that is conducted during the analysis phase and the feature extraction process. While prior literature categorised extraction methods according to data type and analysis methodology (static, dynamic, and hybrid), the focus of this survey is on the technique employed during the feature extraction phase, rather than the data type or analysis approach. Therefore, our survey acts as a threshold between those two phases. In addition to the traditional feature extraction strategies previously present in the literature, modern approaches are implemented. Furthermore, this review discusses and compares those methodologies, as well as the modern methods.

3. DATA BREACHES ON MALWARE ANALYSIS

On the other hand, a regression model is helpful for statistically predicting cyberattacks or anticipating the impact of an attack, such as worms, viruses, or other forms of harmful software. This can be accomplished by analysing historical data. Techniques of regression have the potential to be useful for the development of quantitative security models [27], such as phishing during a particular time period or network packet parameters.

Linear regression, polynomial regression, Ridge regression, and Lasso regression, as well as other well-known regression techniques [28], can be utilised to build a quantitative security model in machine learning. For instance, the authors in [29] employ a linear regression method to trace the roots of a cyberattack, and the authors in [30] employ a multiple regression method to link individual traits to specific online behaviours. Both [29] and [30] give illustrations. Regression regularisation methods like Lasso, Ridge, and ElasticNet are able to improve the analysis of security breaches. This is because the dimensionality of the data pertaining to cyber security is extremely high.

In [31], the authors explore both the predictability of returns on the most prominent cryptocurrencies and the profitability of trading strategies supported by ML methodologies. They take a deep dive into both of these issues. The researchers employ regression models to predict profits on the cryptocurrency being studied. Binary trading recommendations of either "buy" or "sell" are generated using classification models. The expected profits of the dependant variable are another application of regression models.

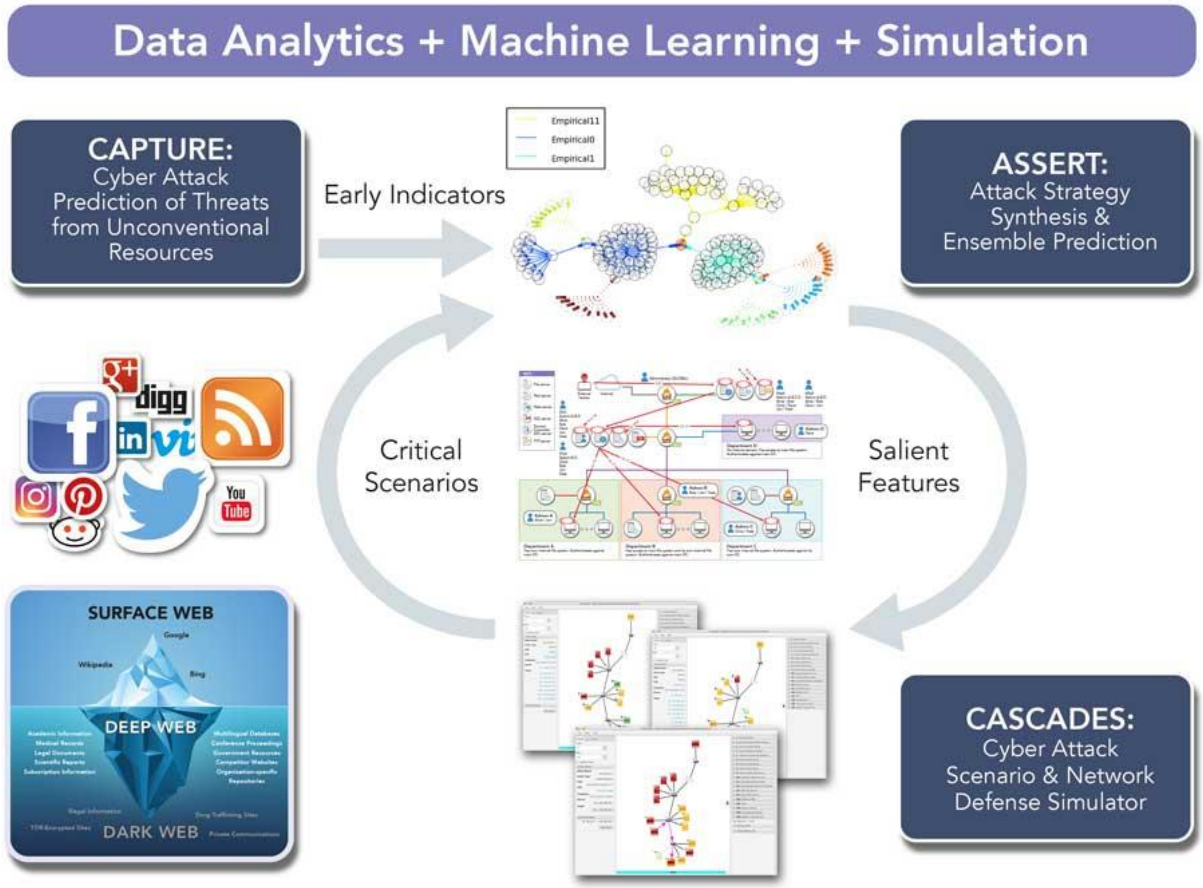
Data Analytics + Machine Learning + Simulation
CAPTURE:
Cyber Attack Prediction of Threats from Unconventional Resources
Early Indicators
ASSERT:
Attack Strategy Synthesis & Ensemble Prediction
Critical Scenarios
Salient Features
SURFACE WEB
Google
Wikipedia
Bing
DEEP WEB
DARK WEB
CASCADES:
Cyber Attack Scenario & Network Defense Simulator

Fig 2: An example of cyberattack progressing into the network through a plethora of techniques.

3.1 Combining approaches to cyber attacks prevention

The National Science Foundation (NSF) in the United States is providing funding for both of the projects that Professor Yang's group is working on. Attack Strategy Synthesis and Ensemble Predictions of Threats (ASSERT) is the first initiative that uses observable malicious behaviour on networks to build predictions about future attacks. Dr. Yang thinks it's possible to prevent such an occurrence from happening by using ASSERT to create a plan to differentiate between ongoing harmful actions and respond to the coming critical risk. ASSERT is able to recognise and anticipate imminent threats by making use of data sources that can be discovered on networks. These data sources include warnings from intrusion detection systems and system logs. In order to produce and improve hypothetical assaults on computer networks, this work blends adaptive Bayesian learning with Clustering and divergence metrics from Information Theory. Two different deep learning algorithms that focus on processing sequence data are then fed the same data in order to categorise cyber threats. Two such strategies are the Long-Short-Term-Memory (LSTM) Network and the Generative Adversarial Network (GAN).

When it comes to learning from very big security datasets, deep learning (DL), a subset of machine learning that evolved from the Artificial Neural Network (ANN), excels above and beyond traditional machine learning techniques. From the ANN came deep learning (DL). The artificial neural network (ANN) combines numerous processing layers into a single network for data-driven learning. Input, intermediate, and output layers are all part of this processing stack. Due to its knowledge-capture nature in deep architecture, deep learning techniques are classified as hierarchical learning methods [32]. These techniques have the potential to learn from cybersecurity data, such as intrusion detection, over several levels.

Thus, we may deduce that using relevant data in the field of cyber security, classification techniques can be employed to construct the prediction and classification model [33]. When considering the security features and the outcome, regression approaches are typically employed to ascertain the model's efficacy [34]. This is achieved by the identification of the strength of predictors, the identification of time-series causes, or the identification of the influence of the relations. To achieve better outcomes in a specific problem domain, it may be helpful to study how to create an effective classification and regression algorithm or data-driven model that utilises relevant cyber data.

4. MALWARE ANALYSIS

An extensive probabilistic risk analysis paradigm was presented by Paté-Cornell et al. (2018) [35] for the specification of cybersecurity in an organisation. Distributions of losses from cyberattacks are shown, both with and without considering preventative actions, to help with risk management decisions based on past and future occurrences. Data for this study came from the Common Vulnerability and Exposures database as well as a confidential database of cyberattacks on a significant organisation in the United States. Sheehan et al. (2021) [36] put forward a novel conceptual framework for the classification and evaluation of cyber risks.

This methodology demonstrated the significance of both preventative and corrective measures in lowering a company's exposure to cyber risk and assessing the severity of that risk. Mukhopadhyay et al. (2019) [37] came up with an additional strategy for assessing and mitigating cyber risk in their research. They employed generalised linear models to approximate loss data linked with each malicious attack, forecasted the security technology needed to mitigate attack risks, and estimated attack probabilities. The insured was advised to obtain cyber insurance after an assessment of the estimated damage from cyberattacks and the net premium that would need to be imposed by a cyber insurer were completed. The information gathered was based on the CSI-FBI poll's findings (1997-2010).

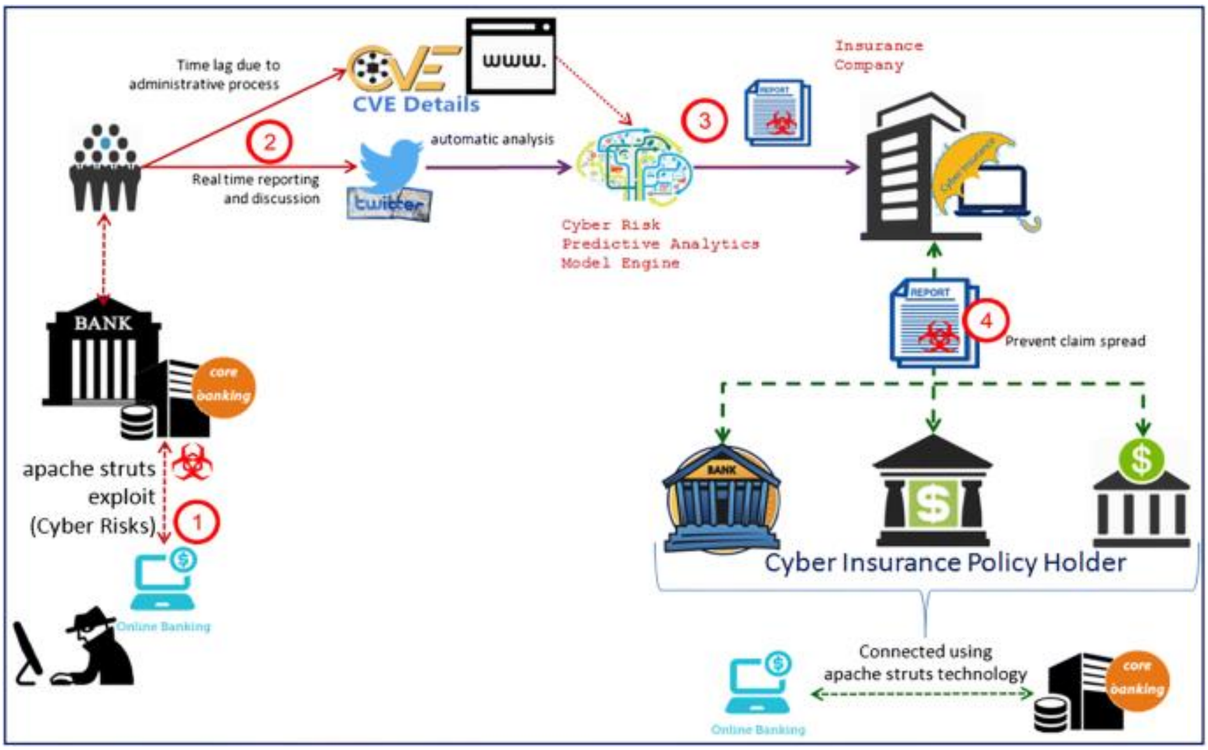
Time lag due to administrative process
www.
CVE Details
2
Real time reporting and discussion
twitter
automatic analysis
Cyber Risk Predictive Analytics Model Engine
3
REPORT
Insurance Company
Cyber Insurance
BANK
core banking
apache struts exploit (Cyber Risks)
1
Online Banking
REPORT
4
Prevent claim spread
BANK
Cyber Insurance Policy Holder
Connected using apache struts technology
Online Banking
core banking

Fig 3: Predictive model implementation scenario illustration.

Eling (2020) [38] reviewed the literature on cyber risk and cyber insurance to demonstrate the dearth of knowledge on the subject. Information about the incidence, severity, and reliance structure of cyber hazards that was readily available was disregarded. Additionally, research gaps related to cyber risk were identified.

The research conducted by Sornette et al. (2013) [39] is yet another example of the gathering of information about the results of cyberattacks. To develop a prediction system that can detect trigger events and prospective accident scenarios, as well as estimate the severity and frequency of the accidents they could produce, they employ a database of newspaper articles, press reports, and other kinds of media.

Using the LexisNexis database, Arcuri et al. (2020) [40] collected a unique sample of global cyberattacks reported by the media. Arcuri et al.'s methodology was strikingly similar to the one used here. The findings here are then applied by Fang et al. (2021) [41] in the areas of dynamic communication and cyber risk perception. Information on the many cyber exchanges between competing governments was compiled by Valeriano and Maness (2014) [42] to construct a database of cyber occurrences and disputes.

Levi (2017) [43] examined international statistics to evaluate the present condition of economic cybercrime and the impact of these numbers on criminal legislation. The characteristics of businesses that are associated to cyber risk awareness were explored by Pooser et al. (2018) [44], who followed the phenomenon from 2006 to 2015. The authors conducted their analysis using data collected from numerous cyber insurance providers. Researchers Walker-Roberts et al. (2020) [45] used the VERIS Community Database to estimate the full extent of damage that could result from a cyber-security breach in a physically and digitally integrated environment.

Bakdash et al. (2018) [46] used US Department of Defence datasets to foresee malware-based cyberattacks. Prediction was used to help do this. This database contains weekly totals of cyber events collected over the course of about seven years. By utilising spatial-temporal analysis, Fan et al. (2018) [47] established an alternative method of prediction for enhancing integrated cybersecurity. As part of their presentation, they offered this technique. Ashtiani and Azgomi (2014) [48] suggested a high-level architecture-based framework for the distributed simulation of cyberattacks. This theoretical structure was also applicable to the study of forecasting. Kirubavathi and Anitha (2016) [49] proposed a method for detecting botnets that would work with a wide variety of botnet architectures. Machine learning and the analysis of network traffic behaviour form the basis of this method.

For network intrusion detection, Dwivedi et al. (2021) [50] presented a multi-parallel adaptive technique that employs an adaptation process among a set of swarms. Recently, AlEroud and Karabatis (2018) [51] released a technique for automatically recognising and searching potential semantic linkages between various sorts of suspicious behaviour obtained from network flows. We were able to do this by considering the material in its proper setting.

CONCLUSION

Malware and other malicious applications have the potential to do catastrophic harm to not only computer systems, but also data centres, web apps, and mobile applications for a wide variety of businesses, most notably those in the healthcare and finance industries. A significant obstacle that brings us to the idea of malware detection and prevention is the need to protect the data of stakeholders from the prying eyes of criminal organisations. Our lives are rapidly getting more digitalized as a result of the rapid progress of technological capabilities. People now live in a cyberworld, in which all information and data are stored digitally and can be accessed online. Whether it be for work or school, banking, shopping, or other errands, almost everything can now be accomplished online. We found that many of the datasets are used for different technical aspects of cybersecurity, and that they are mostly associated with machine learning and intrusion detection. There was a lack of diversity in the preexisting cyber threat datasets. It is extremely difficult for cyber insurance stakeholders to effectively estimate and understand cyber risk due to the ever-changing nature of cyber risk and the absence of previous data. This study's findings could offer a fresh perspective on cyber dangers because they enable the consolidation and classification of cybersecurity databases. This might therefore pave the way for standardised terminology to be developed for cyber legislation. These datasets could be used by businesses that include cybersecurity and cyber risk in their risk management to assess their own cyber posture and the effectiveness of their current protections. Improve your risk awareness and business practises with this publication's thorough overview of peer-reviewed and other publicly available statistics in the subject of cyber risk and cybersecurity.

REFERENCES

1. Nikam, U.V.; Deshmuh, V.M. Performance evaluation of machine learning classifiers in malware detection. In Proceedings of the 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 23–24 April 2022; pp. 1–5. [**Google Scholar**] [**CrossRef**]
2. Akhtar, M.S.; Feng, T. IOTA based anomaly detection machine learning in mobile sensing. *EAI Endorsed Trans. Create. Tech.* **2022**, *9*, 172814. [**Google Scholar**] [**CrossRef**]
3. Sethi, K.; Kumar, R.; Sethi, L.; Bera, P.; Patra, P.K. A novel machine learning based malware detection and classification framework. In Proceedings of the 2019 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), Oxford, UK, 3–4 June 2019; pp. 1–13. [**Google Scholar**]
4. Abdulbasit, A.; Darem, F.A.G.; Al-Hashmi, A.A.; Abawajy, J.H.; Alanazi, S.M.; Al-Rezami, A.Y. An adaptive behavioral-based increamental batch learning malware variants detection model using concept drift detection and sequential deep learning. *IEEE Access* **2021**, *9*, 97180–97196. [**Google Scholar**] [**CrossRef**]
5. Feng, T.; Akhtar, M.S.; Zhang, J. The future of artificial intelligence in cybersecurity: A comprehensive survey. *EAI Endorsed Trans. Create. Tech.* **2021**, *8*, 170285. [**Google Scholar**] [**CrossRef**]
6. Sharma, S.; Krishna, C.R.; Sahay, S.K. Detection of advanced malware by machine learning techniques. In Proceedings of the SoCTA 2017, Jhansi, India, 22–24 December 2017. [**Google Scholar**]
7. Chandrakala, D.; Sait, A.; Kiruthika, J.; Nivetha, R. Detection and classification of malware. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 8–9 October 2021; pp. 1–3. [**Google Scholar**] [**CrossRef**]
8. Zhao, K.; Zhang, D.; Su, X.; Li, W. Fest: A feature extraction and selection tool for android malware detection. In Proceedings of the 2015 IEEE Symposium on Computers and Communication (ISCC), Larnaca, Cyprus, 6–9 July 2015; pp. 714–720. [**Google Scholar**]
9. Akhtar, M.S.; Feng, T. Detection of sleep paralysis by using IoT based device and its relationship between sleep paralysis and sleep quality. *EAI Endorsed Trans. Internet Things* **2022**, *8*, e4. [**Google Scholar**] [**CrossRef**]
10. Gibert, D.; Mateu, C.; Planes, J.; Vicens, R. Using convolutional neural networks for classification of malware represented as images. *J. Comput. Virol. Hacking Tech.* **2019**, *15*, 15–28. [**Google Scholar**] [**CrossRef**][**Green Version**]
11. Firdaus, A.; Anuar, N.B.; Karim, A.; Faizal, M.; Razak, A. Discovering optimal features using static analysis and a genetic search based method for Android malware detection. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 712–736. [**Google Scholar**] [**CrossRef**]
12. Dahl, G.E.; Stokes, J.W.; Deng, L.; Yu, D.; Research, M. Large-scale Malware Classification Using Random Projections And Neural Networks. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing-1988, Vancouver, BC, Canada, 26–31 May 2013; pp. 3422–3426.
13. Tahir, R. A Study on Malware and Malware Detection Techniques. *Int. J. Educ. Manag. Eng.* **2018**, *8*, 20–30. [**Google Scholar**] [**CrossRef**]
14. Gibert, D.; Mateu, C.; Planes, J. The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. *J. Netw. Comput. Appl.* **2020**, *153*, 102526. [**Google Scholar**] [**CrossRef**]
15. Alsmadi, T.; Alqudah, N. A Survey on malware detection techniques. In Proceedings of the 2021 International

Conference on Information Technology (ICIT), Amman, Jordan, 14–15 July 2021; IEEE: New York, NY, USA, 2021; pp. 371–376.

16. Panchariya, H.; Bharkad, S. Comparative Analysis of Feature Extraction Methods of Malware Detection. *IOSR J. Comput. Eng.* **2014**, *16*, 49–54

17. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature selection: A data perspective. *ACM Comput. Surv.* **2017**, *50*, 1–45.

18. El Merabet, H.; Hajraoui, A. A Survey of Malware Detection Techniques based on Machine Learning. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 366–373.

19. Aslan, O.; Samet, R. A Comprehensive Review on Malware Detection Approaches. *IEEE Access* **2020**, *8*, 6249–6271. [**Google Scholar**] [**CrossRef**]

20. AAbusitta, A.; Li, M.Q.; Fung, B.C. Malware classification and composition analysis: A survey of recent developments. *J. Inf. Secur. Appl.* **2021**, *59*, 102828.

21. Caviglione, L.; Choras, M.; Corona, I.; Janicki, A.; Mazurczyk, W.; Pawlicki, M.; Wasielewska, K. Tight Arms Race: Overview of Current Malware Threats and Trends in Their Detection. *IEEE Access* **2021**, *9*, 5371–5396.

22. Oyama, Y. Trends of anti-analysis operations of malwares observed in API call logs. *J. Comput. Virol. Hacking Tech.* **2018**, *14*, 69–85. [**Google Scholar**] [**CrossRef**]

23. Sicato, J.C.S.; Sharma, P.K.; Loia, V.; Park, J.H. Vpnfilter malware analysis on cyber threat in smart home network. *Appl. Sci.* **2019**, *9*, 2763. [**Google Scholar**] [**CrossRef**]

24. Chakkaravarthy, S.S.; Sangeetha, D.; Vaidehi, V. A Survey on malware analysis and mitigation techniques. *Comput. Sci. Rev.* **2019**, *32*, 1–23. [**Google Scholar**] [**CrossRef**]

25. Souri, A.; Hosseini, R. A state-of-the-art survey of malware detection approaches using data mining techniques. *Hum.-Cent. Comput. Inf. Sci.* **2018**, *8*, 3. [**Google Scholar**] [**CrossRef**]

26. Balkrishna, S.; Me, K.; Pratishthan, V.; Shital, M.; Kuber, B. A Survey on Data Mining Methods for Malware Detection. *Int. J. Eng. Res. Gen. Sci.* **2014**, *2*, 672–675.

27. Sarker IH, Hasan Furhad M, Nowrozy Ra (2021) AI-driven cybersecurity: an overview, security intelligence modeling and research directions. SN Comput Sci 2(3):1–18

28. Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2(3):1–21

29. Lalou M, Kheddouci H, Hariri S (2017) Identifying the cyber attack origin with partial observation: a linear regression based approach. In: 2017 IEEE 2nd international workshops on foundations and applications of self* systems (FAS* W). IEEE, pp 329–333

30. Gratian M, Bandi S, Cukier M, Dykstra J, Ginther A (2018) Correlating human traits and cyber security behavior intentions. Comput Secur 73:345–358

31. Sebastiao H, Godinho P (2021) Forecasting and trading cryptocurrencies with machine learning under changing market conditions. Financ Innov 7(1):1–30

32. Amine FM, Leandros M, Sotiris M, Helge J (2020) Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. J Inf Secur Appl 50:102419

33. Sarker IH (2021) Cyberlearning: effectiveness analysis of machine learning security modeling to detect cyber-anomalies and multi-attacks. Internet Things 14:100393

34. Jaganathan V, Cherurveettil P, Muthu SP (2015) Using a prediction model to manage cyber security threats. Sci World J

35. Paté-Cornell, M.E., M. Kuypers, M. Smith, and P. Keller. 2018. Cyber risk management for critical infrastructure: A risk analysis model and three case studies. *Risk Analysis* 38 (2): 226–241. https://doi.org/10.1111/risa.12844.

36. Sheehan, B., F. Murphy, A.N. Kia, and R. Kiely. 2021. A quantitative bow-tie cyber risk classification and assessment framework. *Journal of Risk Research* 24 (12): 1619–1638.

37. Mukhopadhyay, A., S. Chatterjee, K.K. Bagchi, P.J. Kirs, and G.K. Shukla. 2019. Cyber Risk Assessment and Mitigation (CRAM) framework using Logit and Probit models for cyber insurance. *Information Systems Frontiers* 21 (5): 997–1018. https://doi.org/10.1007/s10796-017-9808-5.

38. Eling, M. 2020. Cyber risk research in business and actuarial science. *European Actuarial Journal* 10 (2): 303–333.

39. Sornette, D., T. Maillart, and W. Kröger. 2013. Exploring the limits of safety analysis in complex technological systems. *International Journal of Disaster Risk Reduction* 6: 59–66. https://doi.org/10.1016/j.ijdrr.2013.04.002.

40. Arcuri, M.C., L.Z. Gai, F. Ielasi, and E. Ventisette. 2020. Cyber attacks on hospitality sector: Stock market reaction. *Journal of Hospitality and Tourism Technology* 11 (2): 277–290. https://doi.org/10.1108/jhtt-05-2019-0080.

41. Fang, Z.J., M.C. Xu, S.H. Xu, and T.Z. Hu. 2021. A framework for predicting data breach risk: Leveraging dependence to cope with sparsity. *IEEE Transactions on Information Forensics and Security* 16: 2186–2201. https://doi.org/10.1109/tifs.2021.3051804.

42. Valeriano, B., and R.C. Maness. 2014. The dynamics of cyber conflict between rival antagonists, 2001–11. *Journal of Peace Research* 51 (3): 347–360. https://doi.org/10.1177/0022343313518940.

43. Levi, M. 2017. Assessing the trends, scale and nature of economic cybercrimes: overview and Issues: In Cybercrimes, cybercriminals and their policing, in crime, law and social change. *Crime, Law and Social Change* 67 (1): 3–20. https://doi.org/10.1007/s10611-016-9645-3.

44. Pooser, D.M., M.J. Browne, and O. Arkhangelska. 2018. Growth in the perception of cyber risk: evidence from U.S. P&C Insurers. *The Geneva Papers on Risk and Insurance—Issues and Practice* 43 (2): 208–223. https://doi.org/10.1057/s41288-017-0077-9.

45. Walker-Roberts, S., M. Hammoudeh, O. Aldabbas, M. Aydin, and A. Dehghantanha. 2020. Threats on the horizon: Understanding security threats in the era of cyber-physical systems. *Journal of Supercomputing* 76 (4): 2643–2664. https://doi.org/10.1007/s11227-019-03028-9.

46. Bakdash, J.Z., S. Hutchinson, E.G. Zaroukian, L.R. Marusich, S. Thirumuruganathan, C. Sample, B. Hoffman, and G. Das. 2018. Malware in the future forecasting of analyst detection of cyber events. *Journal of Cybersecurity*. https://doi.org/10.1093/cybsec/tyy007.

47. Fan, Z.J., Z.P. Tan, C.X. Tan, and X. Li. 2018. An improved integrated prediction method of cyber security situation based on spatial-time analysis. *Journal of Internet Technology* 19 (6): 1789–1800. https://doi.org/10.3966/160792642018111906015.

48. Ashtiani, M., and M.A. Azgomi. 2014. A distributed simulation framework for modeling cyber attacks and the evaluation of security measures. *Simulation* 90 (9): 1071–1102. https://doi.org/10.1177/0037549714540221.

49. Kirubavathi, G., and R. Anitha. 2016. Botnet detection via mining of traffic flow characteristics. *Computers & Electrical Engineering* 50: 91–101. https://doi.org/10.1016/j.compeleceng.2016.01.012.

50. Dwivedi, S., M. Vardhan, and S. Tripathi. 2021. Multi-parallel adaptive grasshopper optimization technique for detecting anonymous attacks in wireless networks. *Wireless Personal Communications*. https://doi.org/10.1007/s11277-021-08368-5

51. AlEroud, A.F., and G. Karabatis. 2018. Queryable semantics to detect cyber-attacks: A flow-based detection approach. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48 (2): 207–223. https://doi.org/10.1109/TSMC.2016.2600405.