International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 6 No. 22 March - April - 2025

Leveraging Data Analytics for Cybersecurity: Patterns and Trends in Real-World Attacks

Pratima M. Bhalekar[1] & Jatinderkumar R. Saini[2]
[1]*Ashoka Center For Business And Computer Studies, Nashik, India*
[2]*Symbiosis Institute of Computer Studies and Research,*
Symbiosis International (Deemed University), Pune, India.
Corresponding Author – Pratima M. Bhalekar
DOI - 10.5281/zenodo.15296219

Abstract:

The growing incidence and complexity of cyberattacks have made it crucial to adopt data-driven approaches for detecting, analyzing, and mitigating emerging cybersecurity threats. This paper analyses a real-world dataset (cybersecurity_attacks) of 40,000 cybersecurity incidents. This dataset is drawn from various industries and it includes incidents like malware infections, Distributed Denial-of-Service (DDoS) attacks, intrusions and phishing attempts. We have identified patterns in these attack types, measure severity levels, and compute anomaly scores. By using statistical and machine learning techniques, it finds that DDoS attacks as the most common, with malware and phishing also frequently appearing in the dataset. The study also assesses the severity of incidents and based on the factors like operational disruption, data breaches, and mitigation efforts, we categorised them. A distinguished portion involves high-severity attacks and most incidents are of low to medium severity. It highlighting the need for proactive defense mechanisms to prevent escalating threats. Also, to uncover unusual network behaviors, anomaly detection methods are applied, potentially signalling early stages of an attack. These anomaly scores provide valuable insights for prioritizing responses and improving threat detection systems.

The findings of this study offer practical recommendations for enhancing cybersecurity practices, particularly by integrating data-driven insights into defense strategies. Continuous monitoring, real-time analysis, and adaptive security measures are essential for responding effectively to evolving cyber threats. This research underscores the importance of leveraging statistical insights to optimize incident response and reduce the impact of cyberattacks on organizations.

Keywords- Cybersecurity, DDoS, Intrusion Detection, Severity Levels, Anomaly Scores, Network Security, Threat Analysis

Introduction:

The evolving digital era created lots of opportunities to the organizations and individuals to interact with each other. But with this, it also increases various risk of cyberattacks. In recent years, the occurrences and complexities of these attacks have accelerated, which require the development of more effective techniques to detect, prevent and mitigate cybersecurity threats. Cybersecurity breaches has terrible effects like financial loss, data theft etc. Cybercriminals have refined their tactics more and more against traditional security measures as it based on signature-based detection. Predefine rule sets are not so effective in today's world. So, there is need of dynamic, data-driven cybersecurity approach to enhance the accuracy of threat detection and prevention.

Cyber threats come in various forms like DDoS, malware, intrusions etc. with

their distinct characteristics and severity levels. So,the major challenges in cybersecurity area is to identify and categorization of cyberattacks.

Existing research provided significant understanding of individual attack types, but there is gap in studies like multiple attack vectors and their associated behaviours can give comprehensive view of cybersecurity landscape. Maximum existing research focused either on attack or threat detection without considering the broader context of threat environment.

We tried to addressed this gap by conducting an analytical study of cybersecurity attacks using a large and real-world dataset of cybersecurity incidences. Our dataset includes 40,000 records of various cyberattack types and their associated metadata. The aim of this study to explore valuable insights on various cyberattacks, their severity levels, correlations between attack behaviors and severity levels, and discover different attack patterns.

By using statistical techniques like descriptive statistics, frequency distributions and time based trends analysis, we have analysed the data. The dataset consists of various attack types like DDoS, malware and intrusions. These attacks gave critical threats in digital ecosystem. Python is used for data processing and visualization. Our objective was to identify new trends and correlations within the dataset.

The outcome of this research paper contributed in the field of cybersecurity by providing more comprehensive analysis of real world cyber-attacks and their severity levels. Our analysis also focuses on importance of identifying low to medium severity attacks which may be highly severe incidence if not addressed promptly. The study also highlights the importance of considering anomaly scores as reliable metric for threat detection and evaluation of severity. These findings can help cybersecurity professionals to refine their security strategies.

Literature Review:

The increasing occurrence of cyberattacks emphasized the need for robust cybersecurity techniques across various domains. Existing studies have explored different areas of cybersecurity like understanding cyberattack patterns and evaluating the effectiveness of security measures.

Li and Liu (2024) propose a comprehensive review of cyber-attacks and cybersecurity. They explored the evolution of threat vectors and the adoption of emerging technologies such as AI, blockchain, and IoT. Their study recognizes critical challenges in addressing cyber threats and provides insights into future research guidelines to minimized these risks effectively[3]. Hadlington (2018) focuses on the impact of employee attitudes, company size, and age on engagement in risky online behaviors. He conducted study with 515 UK employees and finds a significant negative correlation between positive attitudes toward cybersecurity and risky behavior, particularly among younger employees and those in smaller companies[4].The researchers, Sheth, Bhosale, and Kurupkar (2021) stress the importance of cybersecurity in safeguarding sensitive data across various sectors, including military, governmental, and financial industries, while emphasizing the need for advanced systems to counter evolving cyber threats[5].

Jalali, Siegel, and Madnick (2021) observed decision-making biases in cybersecurity and find that experienced and also inexperienced professionals face challenges in proactive decision-making, especially in understanding delays and predicting cyber events. They advocate for systems-thinking training to improve decision-making in cybersecurity[6]. Reddy and Reddy (2021) explore emerging

technologies and ethical considerations in cybersecurity, highlighting the importance of protecting information among rising cybercrimes[7]. Zwilling et al. (2021) investigate the relationship between cybersecurity awareness, knowledge, and behavior across four countries, revealing that while awareness of cyber threats is widespread, users often rely on basic protective measures, researcher suggesting that there should be more targeted cybersecurity training programs[8].

Patil (2024) highlights the need of cybersecurity in protecting sensitive data across industries against the growing threat landscape and the need for effective strategies to fight against cybercrimes[9]. La Grew (2024) studied cyberattack patterns using honeypots deployed on cloud platforms and residential networks, finding that automated scanning and botnet activity dominate. Author suggesting that foundational security measures should be prioritized over focusing on provider-specific attack patterns[10].

Jamal et al. (2023) provide a review of cybersecurity analysis methods for Cyber-Physical Systems (CPS) using machine learning. They explore AI-driven techniques to address vulnerabilities in CPS and highlight the potential of these methods to provide adaptive, robust security solutions against both internal and external threats[11]. Alghamdi (2021) analyses traditional signature-based detection methods and advocates for behavior analysis to identify rare and anomalous cyber events, offering insights into improving research accuracy and detecting insider threats[12].

Alibasic et al. (2017) explore cybersecurity challenges in smart cities, highlighting the risks posed by ICT dependence and the need for integrated security solutions to mitigate these vulnerabilities[13]. Arend et al. (2020) differentiate between active and passive-risk behaviors and finding that passive-risk behaviors like neglecting to implement protective actions, are significant predictors of cybersecurity intentions and actual behavior[14]. Kennison and Chan-Tin (2020) highlights the impact of personal behaviour and risk-taking behaviors on cybersecurity practices. They emphasizing the important role of individuals in cybersecurity training[15].

Van ’t Hoff-de Goede, van de Weijer, and Leukfeldt (2023) examine self-protective behaviors like password strength and phishing awareness. They have studies cybercrime victimization through a longitudinal study and its effects[16]. Edgar and Manz (2017) provide basic fundamentals of cybersecurity such as vulnerabilities, exploits, malware etc. and also highlights the challenges of securing cyberspace [18]. Bhalekar and Saini (2024) highlighting the ability of graph databases like Neo4j to create relationships between entities and provide insights of data analysis and data visualization in the field of cybersecurity[19].

Bhol et al. (2023) recommend a taxonomy of cybersecurity metrics. They shows the importance of measurable metrics, which help to to assess cybersecurity strength and effectiveness. They used multi-criteria decision-making approach which provide a systematic way to evaluate and manage cybersecurity efforts [20].

Our literature review addresses both human factors and advanced technologies to understand growing cybersecurity challenges. It also shows the importance of data-driven approaches, integrating behavioural analytics etc. to develop more robust cybersecurity systems.

Methodology:

We used cybersecurity dataset (cybersecurity_attacks). This dataset consists of 40,000 entries and 25 columns, focusing on network traffic and cybersecurity related events. Key attributes of this dataset are

network data, indicators, attack information, severity & context and logs. The network data consists of IP address of source and destination, ports, protocol, packet length, packet type, and traffic type. Indicators includes malware indicators, anomaly scores, alerts and warnings. Attack information consists of type of attacks, attack signature, and action taken. Severity & Context stores data like severity levels, user details, network segment, geo-location data etc. Under logs, it has firewall logs, IDS/IPS alerts and log sources.

The objective of the methodology was to reveal patterns in attack types, evaluate severity distributions, and analyze anomalies in network traffic. We used a combination of descriptive statistics, frequency distribution, and time-series trend analysis, to extract meaningful insights from the various cybersecurity incidences. Descriptive statistics was used to summarize the dataset. It focused on attack frequency and the distribution of severity levels. The frequency distribution was used to find the most common attack types, and time-trend analysis helped us to detect attack occurrences over time.

The data pre-processing was conducted using Python. In which was dataset was cleaned, normalized and transformed to ensure accuracy. Visualizations such as histograms and line charts were used to illustrate attack patterns and anomaly scores. Also the correlations between attack types, severity levels and anomaly scores was find using statistical methods. We have applied anomaly detection techniques to explore unusual network behaviors. This methodology helped us to identify trends and attacks patterns in real world cyberattacks dataset, which is helpful in implementing future advanced strategies to prevent network from attacks like malware, intrusion, DDoS, etc.

Results:

1. Descriptive Statistics:

Table 1: Descriptive Statistics of data

Metric	Packet Length	Anomaly Scores
Count	40,000	40,000
Mean	781.45	50.11
Standard Deviation	416.04	28.85
Min	64	0.00
25th Percentile	420	25.15
Median(50%)	782	50.35
75th Percentile	1,143	75.03
Max	1,500	100.00

The analysis finds that packet lengths ranged from 64 to 1,500 bytes, with a mean of 781 bytes and a median of 782 bytes. This distribution shows that the majority of packets fall within standard transmission sizes, which is typical for regular network traffic. On the other hand, Anomaly scores had an average of 50.1, which shows that the most network traffic has moderate anomalies. On the other hand, the significant variance in the scores points to the presence of considerable outliers. It represents critical or irregular network behavior. These outliers with anomaly scores above the average, are mainly important as they highlight potential vulnerabilities or attacks that deviate from normal network traffic patterns. It suggests enhanced investigation and prevention.

2. Attack Type Distribution:

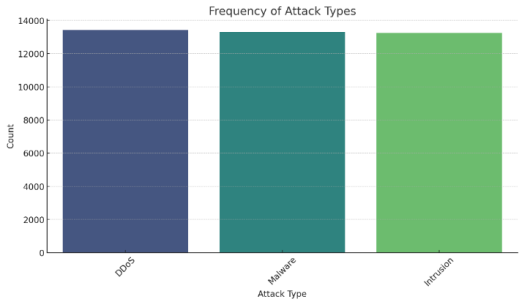

Figure 1 : Attack Type Distribution

The results shows the distribution of three major attack types like Distributed Denial of Service (DDoS), Malware attacks

and intrusions. The results indicate nearly equal distribution of all mentioned attacks. The DDos attacks accounting for 13,428 occurrences which is 33.6% of the total occurrences. Malware attacks followed with 13,307 occurrences which is 33.3% of total occurrences, and intrusions accounting for 13,265 occurrences which is 33.2% of total occurrences. The high frequency of DDoS attacks emphasizes the critical need for robust traffic filtering mechanisms to avoid service interruptions. Malware also remains a persistent threat, which is responsible for exploiting vulnerable endpoints or user behaviors, which could suggest the importance of enhanced endpoint security and user awareness. Also, the high frequency of intrusion occurrences indicates possible weaknesses in network segmentation and access control, emphasizing the need for stronger perimeter defenses strategies and more rigorous access management protocols to prevent unauthorized system access.

3. Severity Level Analysis:

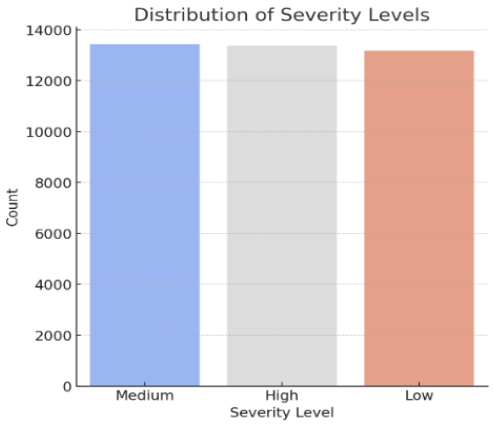

Figure 2 : Distribution of Severity Levels

The Severity Level Analysis analyse the distribution of severity levels. The severity levels of the attacks were distributed into 3 categories that are Low, Medium and High. As per the results, Medium severity accounted for 13,435 attacks (33.6%), High severity for 13,382 attacks (33.4%), and Low severity for 13,183 attacks (32.9%). The analysis shows the balanced distribution which indicates that cyberattacks are not only focused on high-severity incidents but also on medium and low severity attacks that may be unaddressed. Though, medium severity events considered less critical but it could be the reason for significant risks and rising severe issues, if not addressed promptly. This analysis shows the importance of monitoring and responding to all levels of attacks to ensure comprehensive cybersecurity protection.

4. Time-based Trends:

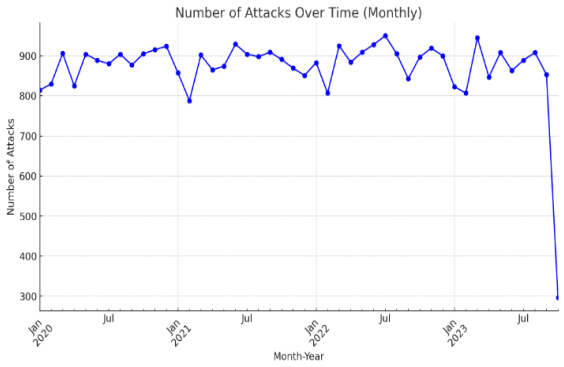

Figure 3 : Number of Attacks over time

The Time-based trends check for any patterns over time. The time-series analysis demonstrated that the number of attacks varied across different months. Spikes in attack frequency may associate with specific events, like phishing attempts during financial quarters. The results highlight the importance of continuous monitoring to detect sequential attack patterns.

Conclusion:

We have explored real world, huge amount of cyber-attacks incidences and done an analysis on it to find key insights on attack patterns, severity levels and anomaly scores. The results show the occurrences of critical cyber-attacks like DDoS, malware and intrusion attacks. The findings show almost equal distribution of attack types which highlights the need of versatile strategies to prevent organisations from critical threats which may include endpoint

security, traffic filtering and secure access control methods. Our severity analysis also shows the need to pay attention not only on high severity attacks but also on low and medium severity attacks. We cannot ignore these attacks as they may cause critical threats if not addressed promptly. It is essential to continuously monitor all levels of cybersecurity threats, for maintaining proactive security environments. The anomaly scores helped in identifying deviations in network behaviour. The significant variance in anomaly scores suggests immediate action on the occurrences of critical threats. The time based analysis help to identify periodic or event driven spikes in attacks occurrences. These insights are crucial for enhancing predictive models and developing targeted defence strategies that detect emerging cyber threats.

Overall, our research supports the need for proactive mechanism that focuses on early threat detection and prevention across different domains. It will help organisations to enhance their ability to protect against the emerging cyber threats.

References:

1. Stallings, W. (2019). *Network Security Essentials: Applications and Standards*. Pearson Education.
2. (cybersecurity_attacks) Dataset
3. Li, Y., & Liu, Q. (2024). A comprehensive review study of cyber-attacks and cyber security: Emerging trends and recent developments. https://doi.org/10.1016/j.egyr.2021.08.126
4. Hadlington, L. (2018). Employees' attitude towards cybersecurity and risky online behaviours: An empirical assessment in the United Kingdom. *International Journal of Cyber Criminology, 12*(1), 269–281. DOI: 10.5281/zenodo.1467909.
5. Sheth, A., Bhosale, S., & Kurupkar, F. (2021). Research paper on cyber security. *Contemporary Research in India, Special Issue: April 2021*. ISSN 2231-2137.
6. Jalali, M. S., Siegel, M., & Madnick, S. (2021). Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. *Journal of Strategic Information Systems*. Retrieved from www.elsevier.com/locate/jsis.
7. Reddy, G. N., & Reddy, G. J. U. (2021). A study of cyber security challenges and its emerging trends on latest technologies.
8. Moti Zwilling, Galit Klien, Dušan Lesjak, Łukasz Wiechetek, Fatih Cetin & Hamdullah Nejat Basim (2020): Cyber Security Awareness, Knowledge and Behavior: A Comparative Study, Journal of Computer Information Systems, DOI:10.1080/08874417.2020.1712269
9. Patil, A. A. (2024). Cyber security challenges and threats. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), 4*(1), 561. DOI: 10.48175/IJARSCT-15082.
10. La Grew, J. (2024). Lack of intentionality: Honeypots show us wandering drones.
11. Jamal, A. A., Majid, A. A. M., Konev, A., Kosachenko, T., & Shelupanov, A. (2023). A review on security analysis of cyber physical systems using Machine learning. *Materials today: proceedings*, *80*, 2302-2306.
12. Al-Ghamdi, M. I. (2021). Effects of knowledge of cyber security on prevention of attacks. *Materials Today: Proceedings*, *10*.
13. Alibasic, A., Al Junaibi, R., Aung, Z., Woon, W. L., & Omar, M. A. (2017). Cybersecurity for smart cities: A brief review. In *Data Analytics for Renewable*

Energy Integration: 4th ECML PKDD Workshop, DARE 2016, Riva del Garda, Italy, September 23, 2016, Revised Selected Papers 4 (pp. 22-30). Springer International Publishing.

14. Arend, I., Shabtai, A., Idan, T., Keinan, R., & Bereby-Meyer, Y. (2020). Passive-and not active-risk tendencies predict cyber security behavior. *Computers & Security*, *97*, 101964.
15. Kennison, S. M., & Chan-Tin, E. (2020). Taking risks with cybersecurity: Using knowledge and personal characteristics to predict self-reported cybersecurity behaviors. *Frontiers in Psychology*, *11*, 546546.
16. van 't Hoff-de Goede, M. S., van de Weijer, S., & Leukfeldt, R. (2023). Explaining cybercrime victimization using a longitudinal population-based survey experiment. Are personal characteristics, online routine activities, and actual self-protective online behavior related to future cybercrime victimization? *Journal of Crime and Justice*, *47*(4), 472–491. https://doi.org/10.1080/0735648X.2023.2222719
17. Lukasik, S. (2020). *Protecting critical infrastructures against cyber-attack*. Routledge.
18. Edgar, T. W., & Manz, D. O. (2017). Science and cyber security. *Research methods for cyber security*, 33-62.
19. P. M. Bhalekar and J. R. Saini, "Comprehensive Exploration of the Role of Graph Databases like Neo4j in Cyber Security," *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, 2024, pp. 1-4, doi: 10.1109/ESCI59607.2024.10497325.
20. Bhol, S. G., Mohanty, J. R., & Pattnaik, P. K. (2023). Taxonomy of cyber security metrics to measure strength of cyber security. *Materials Today: Proceedings*, *80*, 2274-2279.