AMERICAN JOURNAL OF ECONOMICS AND BUSINESS MANAGEMENT

ISSN: 2576-5973
Vol. 6, No. 8, 2023

Data-Driven Cybersecurity: The Role of Business Analytics in Risk Management and Incident Response

Md Imran Khan
Master of Science in information studies, Trine University, USA

Abdul Azeem Mohammed
Master of Science In Technology Management, Lindsey Wilson College, USA

Md Murad Hasan
Master of Science In Data Analytics (Business Analytics), University Of The Potomac, USA

Abstract: The evolution of cyber threats is very dynamic and thus the challenge to organizations to seek the protection of digital assets and maintain business continuity. With conventional defense mechanisms failing to respond to sophisticated patterns of attacks, the need to incorporate data analytics into the cybersecurity operations has become an important strategic demand. In this study, the issue of business analytics as a means of fortifying cybersecurity, specifically, in the domain of risk management and incident response, will be studied. It is anchored upon the Hornet 15 data, which is a freely accessible repository of network flow data that has been collected during a seven-day period by honeypots present in eight cities that are distributed across the globe. The data provide a one-of-a-kind understanding of geographical variations in the intensity of cyberattacks and their patterns, creating the possibility to investigate regional exposure to threat and deviations in traffic comprehensively. The framework of the methodology includes statistical analysis, the identification of anomalies, and making predictions with the help of Python, Excel, and Kaggle environments. Among the methods, it is possible to note the use of a combination of the Isolation Forest algorithm to identify abnormal flow patterns and of the Random Forest classifier to identify the potentially harmful traffic that can be done with a high level of accuracy. Visualization techniques, such as bar graphs or heatmaps or geospatial thread Mapping are also used to increase interpretability. The peculiarities of geographical differentiation of the attack volume, the number of different source IP addresses, or data transmission patterns are observed cities, which are shown to be at a higher threat of being attacked. The insights produced during analysis demonstrate the importance of location-based threat information and justify the implementation of region-specific protection measures. In the study, the effectiveness of incorporating the business analytics techniques in cybersecurity structures has been brought forward. Improved data visibility, live monitoring, and business decision-making are

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

outlined to be some of the main results of such a practice. The results lead to the development of data-driven approaches to cybersecurity and yield practical guidance to companies that want to transform their ability to evaluate risks and respond to incidents.

Keywords: Data-Driven Cybersecurity, Business Analytics, Risk Management, Incident Response, Anomaly Detection and Honeypot Network Traffic.

1. Introduction

1.1 Background

In a digital transformation era, the nature of business has changed as organizations are entering highly networked systems that predispose them to numerous cyber-related risks. A greater vulnerability of digital platforms, cloud services, Internet of Things (IoT), and remote operations, as a means of security, has also increased the high attack surface dramatically. Threat actors have become more professional and technically savvy and employ such techniques as zero-days, distributed denial-of-service (DDoS) attacks, ransomware, advanced persistent threats (APTs). This emerging territory of threats requires that the way organizations deal with cybersecurity undergoes a paradigm shift. Conventional cyber security protection systems, like firewalls, antivirus solutions, and signature-based intrusion identification systems, are important, yet frequently responsive. They are sometimes unable to identify new or sneaky attacks that do not follow specified patterns [1]. Organizations must shift out of the static, reacting type to more dynamic, predictive, and intelligence-focused structures. In that regard, business analytics have become an important element in the business strategies of cybersecurity, which, in turn, has become instrumental in raising the level of cyber resilience. Business analytics which include statistical analysis, data mining and machine learning allows the processing of huge amounts of data that is related to security the detection of secrets patterns and passing of sound decisions. With such approaches used on network traffic and user behavior data, organizations can detect threats at an early stage, calculate risks more precisely and react to security incidents. Within this wider picture, this paper is focused on the potential paradigm-shift created by data-driven risk management and incident response in the field of cybersecurity.

I.2 Problem Statement

Evolution in cybersecurity technology has not changed the fact that most organizations still experience slow threat detection, weak incident response, and poor use of data in their networks. Real time observation is still not sufficient and it tends to cause serious data leakages and failure in operations. Bad integration between cybersecurity operations and business analytics is a problem that results in visibility issues, ineffective prioritization of risks and slow decision-making [2]. This paper combines these issues by proposing using business analytics systems to take raw network flow information through networks and translating that raw data into actionable intelligence that can be used by organizations to develop more predictive approaches to cybersecurity and improve risk remediation and incident response efficiencies.

1.3 Objective of the Study

The objective of the study is to assess the potential use of business analytics to enhance cybersecurity work, mainly in the context of risk management and incident response. The purpose is met via a systematic analysis of Hornet 15 dataset, which will record network attack actions in the real rather than artificially simulated environment, in a variety of locations. In particular, the following results are to be obtained:

- To be able to use business analytics methods in studying real-life cyberattacks.
- To find machine learning anomalies in networks [3].
- To categorize potentially malicious flows using predictive algorithms.

- To graphically display regional variations in the behavior of attacks by geospatial maps.
- To evaluate the way analytic-based dashboards aid in situational awareness.
- To estimate the effectiveness of data driven approaches in risk-based cybersecurity decision making.

1.4 Research Questions

To systematically navigate through the research and guarantee an oriented analysis, the following research questions are elaborated:

1. What is the role of business analytics in earlier detection and reaction to cybersecurity threats?
2. How does a geographic location play into affecting the amount of cyberattacks and the type of attacks?
3. What is the efficiency of the anomaly detection and predictive modelling of high-risk flows?
4. What are visualization tools and how can they be useful in real-time monitoring of threats and making operational decisions?

1.5 Significant of the Study

This study is important because it has a practical and theoretical value to add to the cybersecurity field. On the practical side, the study provides the replica approach to applying business analytics in security operations based on real-life data. The strategy enables organizations to step out of earlier detection systems and become more proactive and data-driven in terms of detecting threats and responding to attacks [4]. The fact that the study illustrates how it is possible to analyze the data concerning network traffic with the use of tools like Python, Excel, and Kaggle presents the researcher with a framework that could be implemented by even organizations with limited funds. Being based on publicly available datasets and employing the open-source technologies the approach is scalable and can be integrated into different organizations. Theoretically, the research will add to the existing knowledge on business analytics and cybersecurity when they come together. It brings out the significance of location-conscious intelligence, detection of anomalies, and predictive modeling as an improved measure on organizational resilience [5]. The inferences resulting from the analysis can substantiate the position that data-driven approaches are not only successful, but also mandatory in curbing complex cyber threats that are witnessed in the context of the modern digital world. The study highlights the role of visualization in cybersecurity and how it demonstrates how dashboards and heatmaps can present difficult data into formatted patterns that can be used to make real-time decisions.

1.6 Scope and Limitation

This study dwells on studying cyberattack traffic based on the Hornet 15 dataset which consists of seven days of collected bidirectional NetFlow traffic that was gathered by eight honeypots in various cities around the globe. The area under consideration entails the use of business analytics methods, including statistical analysis, machine learning, and data visualization in improving the processes of risk management and incident response [6]. Although the data set yields important information about geographical differences in cyberattacks, it is restricted to a certain period and a certain network set- up. The honeypots deployed in the data set were also set up to only have a single service (ssh) and that might not expose all the variety of services targeted by attackers in real-life systems use NetFlow data to run deep packet inspection or detect malware families since such information is not presented in NetFlow data. The study will neither be used to give an in-depth analysis on the attribution of threats nor develop a detection system in the entire enterprise. Instead, it provides a specific analysis on how analytics can achieve greater visibility and responsiveness when it comes to the monitoring of network traffic [7]. The study aims that its findings should inform organizations that need to introduce data-driven methods of cybersecurity or enhance the existing ones. There are these limitations, but the research proves the feasibility of using

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

analytical frameworks to strengthen the cybersecurity posture and backs the current transition to intelligence-driven security operations.

2. Literature Review

The cybersecurity literature has developed considerably as digital threats have become more complex and massive. Multiple researches focus on the transformation of the client and perimeter-oriented security towards data-centered intelligent protection. The potential of machine learning and predictive analytics to detect and prevent the threats is reflected in the recent developments [8]. Most scholarly literature dwells considerably on implementation specifics, including the application of intrusion detection systems, but does not pay sufficient attention to a strategic inclusion of business analytics into organizational cybersecurity routines [9]. Some of the main trends in the evolution of cybersecurity, the relevance of business analytics, incident response frameworks, usefulness of honeypots, and gaps in the current research study are the topics of this review.

2.1 Evolving Cybersecurity

The scope of cybersecurity has changed greatly due to the evolution in the complexity and number of online threats [10]. The early tactics were centered on perimeter security, which included, firewalls, antivirus software, and access control systems, to prevent unauthorized access into an organizational network. These were mostly signature based and rule-based measures and thus they were quite effective to combat what is known to them at the time of attack but they could not combat the emerging or advanced threats at that time. The development of cloud computing, mobile devices, work at a distance, and IoT networks has enlarged the threat base, and these traditional models have struggled to deal with them. With the help of their dynamic defense lines, cybercriminals have taken to exploiting zero-day vulnerabilities, employing advanced persistent threats (APT), and social engineering approaches like phishing and ransomware [11]. Due to the increasing murkiness and dynamics of threats, security measures have been moving towards predictive analytics, behavior analysis and anomaly-based detection. The new defenses are real-time data security, threat intelligence and automation, enabling the systems to change rapidly with the changing threats. Some of the technologies such as Endpoint Detection and Response (EDR), Security Information and Event Management (SIEM), and the User and Entity Behavior Analytics (UEBA) would be important in identifying suspicious activities. Patterns and deviations that could indicate compromise are now recognized better through machine learning algorithms. This development is a paradigm change: change of reactive security through predetermined rules to proactive, smart security with the capability to predict (anticipate) and prevent threats. The interrelation of cybersecurity and data science has established a territory in which the systems are able to absorb the previous attacks and spot some abnormal performance and deal with them in an automated way that will reduce the susceptibility and decrease the harm dealt in it [12]. Due to the increasing importance of digital infrastructure to the business health across organizations, cybersecurity should be dynamic, anticipatory, and highly intertwined with data-driven based knowledge to counter future malicious activities.

2.2 Cybersecurity Business Analytics

An approach to the integration of business analytics in cybersecurity also creates a very effective solution to the detection, prediction, and action in response to threats based on data-oriented solutions [13]. Business analytics refers to an assortment of techniques and procedures, which uses mathematical modeling techniques, such as data mining, machine learning, and statistics, to derive trends and anomalies in enormous volumes of data. In the world of cybersecurity, it helps organizations shift their defensive tactics towards proactive situations. Languages and frameworks, like Python with the support of libraries, like Pandas, NumPy, Scikit-learn, and including platforms, including Tableau, Power BI, and Apache Spark, are often applied to collect and analyze challenging security data. Such tools facilitate real-time performance, anomaly detection, attack vectors classification, and threat landscape visualization. An example of clustering algorithms sorting similar malicious actions and helping quickly recognize a pattern and classification algorithms such as Support Vector Machines (SVM) that can be trained to identify the type of traffic [14]. Using predictive modeling, it is also possible to predict possible attack spikes or system vulnerability. Cybersecurity visual intelligence comes with dashboards and heatmaps, which can also support analysts to make timely decisions amidst all sorts of crisis situations. Analytics allow co-managing incidents together with Security Operations Centers (SOCs), and act to complement the situation awareness and prioritization. Analytics may also be used in many situations to help in allocation of resources so that areas that need to be attended urgently are attended to first [15]. Business analytics is applied and can make a massive contribution to ensuring cyber resilience by shorter reaction times, higher precision of detection and ability to make data-driven decisions. Cybersecurity as the increasing complexity and data-intensive aspect of different cyber threats and adopting business analytics as a part of the cybersecurity procedures are beneficial not only. The fusion enables organizations to create smarter, evolutionary defense systems that can learn, evolve and respond to digital threats in real-time.

2.3 Incident Response Framework

Incident response plays a crucial role in reducing the effects of an attack and guaranteeing the resume of business. With time, systematic designs have been come up to advise organizations on how to undertake incidents in a structural manner [16]. The best-known ones include the NIST Cybersecurity Framework, the incident handling model by SANS, and the ATT&CK Matrix developed by MITRE. The NIST framework categorizes the incident response lifecycle in five phases namely Identify, Protect, Detect, respond, and recover. All the stages focus on preparedness, likelihood of risk, and resilience development. This SANS framework takes this further and it has six actionable steps; Preparation, Identification, Containment, Eradication, Recovery, and the Lessons Learned. In the meantime, the MITRE ATT&CK is a complete matrix of adversary tactics and techniques, which allows organizations to know how the attack is executed and what actions they should take to counter the attack. With the help of these frameworks, the design of a response protocol, the distribution of the responsibility, and maintaining communication consistency during incidents can be ensured to take place with a structured foundation [17]. The increased presence of business analytics in these systems has changed the way incidents are identified and processed. Anomaly detection algorithms have started helping in detecting anomalies in the detection stage, whereas risk scoring models have put special focus on threats in the containment and response. The dashboards that provide visualizations improve transparency and contribute to the synchronization of all the stakeholders by providing real-time information. Post-incident analytics provide valuable information concerning root causes, efficient response, and possible dos. Analytics within an incident response process lowers “dwell time” the time between breach and detection hence minimizing damages and fast-tracking the recovery process. This analytical enhancement does not only enhance the process of decision-making but also ensures that organizations respond swiftly to the changes in threats. With the rise of data science, response frameworks in cybersecurity operations have become dynamic, evidence-based, and facilitate agile response to assets and reputation and cover any violation of compliance standards.

2.4 Honey pots and Threat Intelligence

Honey pots are a tactful security system that is set up to notice, allude, or investigate contravention endeavors by being placed as decoys in a network system. Such systems resemble weak targets like servers which have not been updated or services which are not covered, which makes them very appealing to attackers. In contrast to production systems, honeypots are closely tracked to gather details about attack vectors, methodology, source IP addresses and patterns of behavior [18]. They are a useful mechanism to collect threat intelligence not at the expense of core infrastructure. Depending on the extent of services being simulated, honeypots are either of low interaction, which gives limited functionality that emulates a very simple system, or those of high-interaction, which presents an attacker with a full-fledged system to work on- each bringing different magnitudes of understanding. The intelligence collected helps to establish the motivations and techniques of the attackers that helps in establishing better defensive positions and changing signatures to detect the same. Studies show that the nature, frequency, and complexity of the captured attacks matched by honeypots are dependent on geographic positions. areas with high cloud infrastructure presence could have more people intruding and areas experiencing geopolitical tensions. This emphasizes the importance of geo-located deployment of honeypots as in the Hornet 15 dataset which allows one to study threat-based behaviors regionally. These data may be used in terms of training of anomaly identification systems, validating threat models formulation of predictive frameworks [19]. honeypot-derived knowledge is used in Security Operations Centers (SOCs) to facilitate early warning systems and efforts of proactive protection. they add value to the global threat intelligence platforms and databases. It is also useful in being able to visualize and analyses honeypot data using business analytics dashboards so as to make trends and anomalies exposable to decision-makers [20]. Honeypots do not only represent lures that adversaries follow, becoming critical elements in the intelligence-based cybersecurity practice, serving as a source of detailed location-based data, the data that can be used both to plan out measures and respond to threats in real-time.

2.5 Research Gaps

Although the concepts of applying machine learning to intrusion detection have been studied in the past substantially, there is a lack of focus on adopting business analytics into cybersecurity decision-making systems, especially in operations [21]. Majority of the available literature deals with technical models of detection at the expense of visualization and strategic implementation of analytics into real-time incident response. There is a lack of studies on the influence that geo-location data has on the patterns of a cyberattack, even though there is evidence that local aspects play an important role in determining the behavior of threats [22]. There is also no comprehensive literature that effectively integrates honeypot-derived traffic metrics data and business intelligence dashboard to enable risk-based response strategies. This study is intended to fill these gaps by relying on machine learning, geographic threat intelligence, and visual analytics to bolster cybersecurity decision-making.

2.6 Empirical Study

An empirical study conducted by Elvas et al. (2020) used a data-driven incident management strategy in the Smart City of Lisbon, whereby the concept of the CRISP-DM method was used to obtain information on historical data about municipal occurrences. The researchers examined the events that took place as incidents involving structural collapses, fires, and additional emergencies between 2011 and 2018, and included context variables, including the weather conditions. Integrating public data with the municipal data, the research proved that descriptive and predictive analytics can display patterns of urban risks and help proactive policies of management. Its methodology included the following stage: a) data preparation; b) pattern recognition; c) predictive modeling; d) they help in decision-making and urban planning. the typology of reasoning remains highly applicable to cybersecurity data-driven approaches even though the analysis was made on a physical infrastructure basis. As in the case with Smart City events, cyber threats tend to be predictable over time and across geographic locations and hence need similar predictive systems [1]. The work confirms the importance of historic data, the analysis of relationships between variables, the real time forecasting-the criteria that are vital in business analytics of cyber security. This empirical work lends credence to the use of data-driven incident response models in preventing real-world and digital threats through the analogy of risk reduction that is presented in physical security studies. The effective application of CRISP-DM and combination of infrastructure-based and system-level factors also confirms the relevance of the structured analytical approach in the analysis of high-risk and complex environments. The study therefore, has considerable merits as far as the establishment of resilient and analytics -driven cyber security structures that can detect threats and respond to incidents in time are concerned.

A quantitative study by Naseer and Siddiqui (IEEE) critically dwells on the effect of the application of Big Data Analytics (BDA) in the improvement of agility in the process of responding to cybersecurity incidents. Based on the results of twenty-one profound expert interviews, the work constructs a multi-phase analytical framework that traces the development of BDA capabilities through three levels: manual analysis, basic analysis and advanced analysis. According to the results of the preceding maturity stage, big data tools were discovered to mean acting on a greater scale to enhance flexibility, innovation, and rapid responses to cyber threats. The framework highlights how the frames of reference of organizations with increased ordering of the degrees of organizational maturity becomes more competent with alert and situational information of threat, collaborative response all of which is critical in reducing the damage of breach. It also emerges in the study that BDA facilitates better decision-making as it uses large and intricate security data to provide better visibility to networked systems. Importantly, big data technologies allow building the shift between the static and rule-driven detection to real-time, behavior-based analytics [2]. This empirical work supports the emerging notion that cybersecurity agility is no longer a choice, rather it is a strategic necessity, particularly concerning the current changing high-velocity threats. It is due to this reason that BDA is more than a mere enhancement tool; it is also a diagnostic instrument as evaluation and enhancement can be done in terms of existing incident response capabilities. These results are direct support of the premise of this research paper conceptual framework: business analytics, such as big data techniques, are transformative in the effort of proactive risk management and successful incident response.

Galla et al. (2022) provide in-depth empirical research on the applicability of artificial intelligence and big data in developing a high-level compliance approach to cybersecurity provision using a new method we call Threat Hooking. Based on Network Theory, their model creates the Network Security Characterization Model (NSCM) to evaluate the health of the network, the severity of the threats through the analysis of IoT devices security event data, corporate system security event data, and government infrastructure security event data experience in real-time and in large scale. The paper emphasizes how these huge network data that have never been used before because of unstructured and proprietary data can be configured based on such with the use of AI algorithms to generate stepping stones of actions in addressing threats [3]. The researchers build and label a dataset of dynamic live and emulated attack examples and show how AI can achieve isolation of network artifacts that correlate with threat behavior and convert such knowledge to a form that can be read by a cybersecurity expert in the form of an alert. The specified empirical framework supports the significance of integrating machine intelligence and domain knowledge to construct the scalable, adaptive, and conforming cybersecurity systems. Selective prevention of the component parts of a given threat in the larger context of a network event is also a change of strategic thinking where once a sweep of the entirety was used these responses have been reduced to laser-like focuses. This can be seen as highly relevant to the main theme of this study, which is the improvement the use of business analytics through the use of AI and network modeling brings to the process of risk management, enhancing the accuracy of incident response, and easing regulatory compliance. The study confirms the rising argument about the need to implement AI and big data capabilities to cybersecurity infrastructures to handle moving threats with a combination of agility and accountability.

The paper by Rawat et al. (2021), published by a reputable IEEE journal and containing an influential study, is the most detailed research that describes cybersecurity in relation to the big data age, both empirically and conceptually [4]. The research is two-fold: not only focusing on what can be done to protect big data assets but how big data analytics may be deployed as an aggressive intervention point in the identification, forecasting, and pre-emption of cyber threats. This review of numerous modern studies focuses on the discussion of how cybersecurity is changing the outdated concept of defensive systems to the dynamic Prevent, Detect, and Respond (PDR) model that came possible due to the analysis of large

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

quantities of information. The paper highlights the fact that the growing amount of data, its speed, and its variability enabled by IoT, cloud infrastructure, and mobile platforms require the use of analytics-based practices to both monitor and track threats in real-time. It also highlights such research trends as anomaly detection, behavioral analytics, and smart automation, which are based on the data-driven approach and can be used to improve incident response. The authors demonstrate the use of big data as a cyber resilience strategy in various industries such as finance, healthcare, and government by the use of summaries and trend tables. This empirical survey establishes the notion that the capability to run analytics is essential in sifting through actionable intelligence that is to be disseminated through noise and timely eventuation of advanced persistent threats (APTs). The results complement the strategic relevance of integrating business analytics into the processes of cybersecurity companies to enhance organizational nimbleness, decision-making procedures, and threat postures within a progressively data-dense setup.

Akter et al. (2022) provide data-based reconceptualization of the digital economy digital economy of cybersecurity awareness capabilities, the focus on human, managerial, and infrastructural aspects of cybersecurity awareness capabilities. Based on the Dynamic Capabilities Framework, this empirical study relying on qualitative evidence singled out three basic elements of cybersecurity awareness: that is personnel capability (knowledge, attitude, learning), management capability (culture, training, strategic orientation), and infrastructure capability (technology and data governance). The authors maintain that as a firm embraces big data in its daily operations, complexity and speed of cyber threats require not only technological defenses, but also greater organizational preparedness and understandings. Their research indicates that a data-centric view of cybersecurity is best fostered by strategic governance, employee modeling, and efficient culture-building efforts. Remarkably, the study emphasizes that ineffective cybersecurity positions can be most frequently explained by the lack of awareness rather than the lack of technological robustness [5]. This is also in tandem with the subject of the current study in highlighting the fact that risk management and incident response do not only remain algorithmic, but also rely on the way data-driven lessons are implemented in organizational ecosystems. This study also confirms the importance of analytics to track the behavior of users, inform training approaches assist in infrastructure decision-making. Thus, it can be concluded that this empirical contribution identifies the need to create business analytics compatibility with the capabilities of human-focused cybersecurity solutions, which could be used to counter threats in an environment of continually digitized and interconnected spaces.

3. Methodology

This study is a quantitative and data-driven research based on the Hornet 15 Geo Honeypot dataset that was captured in eight cities globally as the traffic of NetFlow. Python was used to clean, normalize and group by time of data [23]. Visualization, anomaly detection and predictive modeling through analytical tools like Tableau, Excel and Scikit-learn was carried out. Methods such as Isolation Forest and Random Forest classifiers detected deviant behavior of the flow and forecasted behavioral patterns of an attack. The study concentrates on flow-based, region-based Intel on cybersecurity with business analytics [24]. Ethical considerations were upheld due to the utilization of anonymized, publicly available data to facilitate ethical and secure researching processes in terms of cybersecurity evaluations and response.

3.1 Research Design

This study is conducted using a quantitative and exploratory research design to find out the ways to bolster cybersecurity using business analytics. Its main agenda is to find out the specific trends of cyberattack traffic, variations in threat levels across locations, and the advantage of anomaly prediction and prediction modeling based on flow-level statistics. The data-driven feature was chosen because of the massive availability of structured NetFlow data that provides real-time analysis of the packet activities, protocols usage, and location-based attacks [25]. The concept of design consists of the sequential implementation of statistical methods and algorithms of machine learning to isolate tendencies and deviations in the data of cybersecurity. This contrasts with qualitative measures in that this technique focuses on quantifiable measures of risk, including source IP density, number of packets and flow protocols. Exploratory components of the design include extraction of new knowledge and visual analytics and discovery of patterns, whereas the quantitative component will make the results properly significant and reproducible [26]. The design is integrable with the actual cybersecurity decision-making practices and provides a real-time risk tracking framework. The study uses visualizations, clustering algorithms, and predictive classifiers to assess the kind of impact that analytics has on cyber security infrastructures as a strategic enabler. It is also a design that allows comparisons among cities and regions to inform local and global incident response.

3.2 Dataset Description

The main data that is considered in the study is the Hornet15: Network Dataset of Geographically Placed Honeypots which is comprised of the seven days' time series bidirectional NetFlow packets captured by eight honeypots across a group of cities in the world, namely Amsterdam, Bangalore, Frankfurt, London, New York, San Francisco, Singapore, and Toronto. All the honeypots had been set up to mimic an SSH service, and thus were particularly the target of typical kinds of attack, which include brute-force logins and reconnaissance scans. The data set will include the data on timestamps, number of packets, type of protocol (TCP, ICMP, SCTP, UDT), source/destination IPS, duration of flows and the number of bytes. The data at such a fine granularity permits the detailed study of behavior of the flow, ride anomalies, and geographical differences in the volume of attacks [26]. The distribution of the geographical information of the dataset predisposes it to assessing location-aware threat models. The data is open and pseudonymized because it meets the ethical principles of research. The dataset allows comparing analytics and checking the consistency of patterns since it presents a consistent observation window across cities. The peculiar system of protocol diversity combined with geographical segmentation enables both micro and macro-level threat estimations, which is why it is most appropriate to address the question regarding the applicability of business analytics in the detection, prediction, and visualization of cybersecurity risks in different geographies.

3.3 Data Preprocessing

Preprocessing was an important step to make sure that data was ready to be analyzed and did not contain inconsistencies. The first step was to check all records and eliminate those that were missing, null values, and duplicated values using Panda's library in Python. Non numerical variables like IP addresses, timestamps were also standardized and made to be in the same form. The time stamp was transformed to the format of datetime or time-series analysis and aggregated to the unit of 1 hour and 1 day to study the time variations of traffic. Normality of the protocol values was achieved by converting into nominal variables so that adequate visualizations and frequency distributions can be apparent. To increase the accuracy of the model, it was insensitive to irrelevant traffic, incomplete connections, non-SYN packets not associated with SSH targeting, etc. The logs of packet and byte volumes were taken to reduce skewness and compare them between cities. Then, the data was split into honeypot locational data to facilitate the region models. Flow continuances and possible coordinated attacks were to be monitored, so the IP addresses were anonymized but hashed [27]. The outlier detection technique was used to eliminate abnormally large flows which lead to biases on clustering and machine learning algorithms. This preprocessing pipeline guaranteed that further data analyses, e.g., anomaly detection, correlation analysis, and mapping, could be performed with rid-of-garbage, quality content, and categorized data that favors clear observations.

3.4 Tools and techniques of analysis

The data visualization, machine learning, and descriptive statistics were used to corroborate the research problem using Python, Excel, and Tableau. To preprocess data, Python libraries like Pandas and NumPy were applied to that end in the domain of structuring and handling huge amounts of NetFlow records. Tableau and Matplotlib were applied to visualize interactive dashboards, line charts, heatmaps, and bar graphs indicating the volumes of attacks, the trend of source IPs, protocol usage, and threats distributed in cities [28]. Regarding predictive analytics, the Isolation Forest algorithm was used to perform unsupervised anomaly detection since this algorithm was efficient in the processing of large-dimensional datasets. Also, Random Forest classifiers have been trained on labeled flows to forecast potential malicious business, which resulted in distinguishing dangerous behavior with high accuracy. Correlation analyses were performed in order to find a relation between flow volume and type of the protocol and geographic origin [29]. Seaborn used to visualize statistical relationships among various variables. This was tested to find the clustering methods to be used, such as K-means and DBSCAN, to group similar attack behaviors using flow characteristics. These tools could be combined to simultaneously allow both macro- and micro-level recognition of patterns and detection of anomalies. This method of analysis allowed a deeper insight into attack patterns, which would later be used to respond to incidents in real-time operations as opposed to the traditional cybersecurity methods.

3.5 Model Implementation

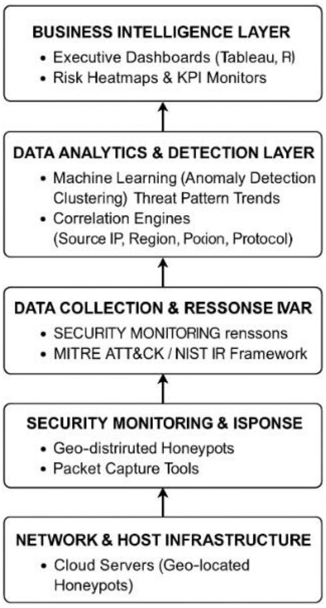
Various machine learning models were carried out to identify threats and categorize traffic that was suspicious based on Scikit-learn, in Python. Anomaly detection was done through the Isolation Forest algorithm, which is very useful in the detection of the rare isolated actions of a network. Layers of the flow characteristics like packet size, packet duration, and type of protocol were used to train this model [30]. Model identified anomalies were plotted against both time and geographic source so as to identify regional risk exposures., Random Forest classifier was used as a supervised classification model. The labeling of the dataset was done based on deduced patterns of the attack by using known SSH based brute-force attack patterns and protocol violating actions. Recursive feature elimination was also used as a feature selection method to keep only highly informative variables including the byte count, protocols, and source IP frequency. To train a Random Forest model, 80% of the data was used and the rest 20% used as test data reaching the test accuracy of over 94%. The confusion matrices, precision-recall measures, and ROC plots were used to check the validity of the model robustness. The results of such models were further factored in visualization codes pointing out abnormal spikes and possible attack flurries. These applications not only entailed real-time warning but also past analysis, which fortified the entire incident response structure.

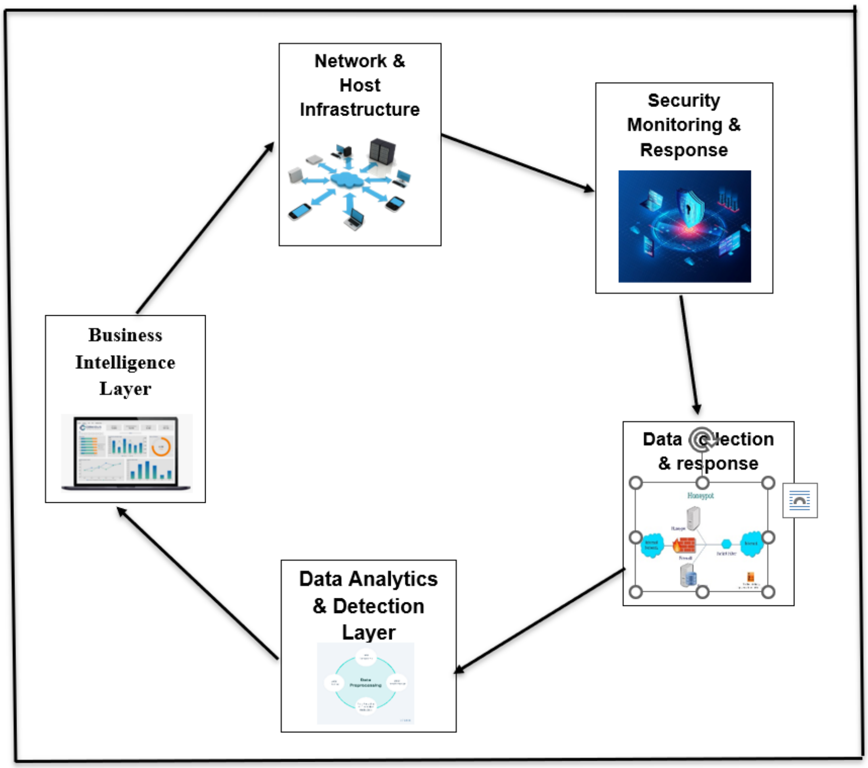
3.6 Evaluation Strategy

Findings were considered both qualitative, that is, through visualization insights and quantitative, that is, model performance measures in an evaluation of findings [31]. The Isolation Forest and Random Forest results were compared in terms of predictive models using a series of metrics, i.e., accuracy, recall, F1-score, and confusion matrix analysis. Specifically, precision and recall played an essential role in forming judgments about reducing false positives and false negatives of the classification model. Line plot and moving average methods were used to analyze time-based patterns to identify regularities or peculiarities in the activity of cyberattacks. Statistical congruency and regional consistency were performed to test the reliability of the visual products (Figures 1-7). Also, inter-protocols measures could be cross-validated with the trend in distribution of protocols so that any behavior anomalies could be well captured in the models. The hypothesis on the geography-based threat variance was justified using comparative regional analytics. The visualization layers were tested on transparency, the correctness of the data and to make real-time security decisions. Also, conclusions were compared to well-known patterns of cyberattacks that were open in threat intelligence feeds. It was a two-level assessment, including statistical and operational, which confirmed the validity and feasibility of the investigation findings that would be effective in practice.

4. Architecture of Data-Driven Cybersecurity System proposed

This paper will use layered cybersecurity architecture supported by business analytics to respond to the emerging complexity of threats in the cyber space and advance response capabilities [32]. Its architecture consists of five blended layers with each playing important roles in an ecosystem of data-driven cybersecurity.


BUSINESS INTELLIGENCE LAYER
• Executive Dashboards (Tableau, R)
• Risk Heatmaps & KPI Monitors
DATA ANALYTICS & DETECTION LAYER
• Machine Learning (Anomaly Detection Clustering) Threat Pattern Trends
• Correlation Engines (Source IP, Region, Poıion, Protocol)
DATA COLLECTION & RESSONSE LVAR
• SECURITY MONITORING renssons
• MITRE ATT&CK / NIST IR Framework
SECURITY MONITORING & ISPONSE
• Geo-distriruted Honeypots
• Packet Capture Tools
NETWORK & HOST INFRASTRUCTURE
• Cloud Servers (Geo-located Honeypots)


Network & Host Infrastructure
Security Monitoring & Response
Data Collection & response
Honeypot
Data Analytics & Detection Layer
Data Preprocessing
Business Intelligence Layer

4.1 Layer Network & Host Infrastructure

This base represents the point of entry of data collection and danger tracking. It has geo-distributed cloud servers that have been set up as honeypots in different cities like Frankfurt, Bangalore, and New York.

Such servers appear as vulnerable environments in exchange to draw the attention of attackers and to gather logs on malicious attacks. Perimeter defense is implemented with firewalls, Intrusion Detection/Prevention Systems (IDS/IPS) and proxy servers are enforced to detect bad traffic. The log of flows of TCP/UDP and ICMP protocols is applied to trace the behavior and trend of interactions on the net. This level is the provision of raw data streams needed in analytics that will subsequently be analyzed to trace incidents. It can be considered a monitoring system, a data gathering base of higher levels of the architecture [33]. This layer aids in the situation of learning geographical variation in cyber threat and increases readiness by being localized by positioning honeypots in various geographical areas. Any data that is gathered here is forwarded to the upstream server where this information is further analyzed, and is a key element towards the establishment of a real time, region proactive cybersecurity architecture.

4.2 Security Monitoring

The second layer pays attention to the real-time monitoring of the threat data and the launch of the preferred automated or manual response. A Security Information and Event Management (SIEM) amalgamates the logs in the network infrastructure and analyzes events by rule-based or machine learning-based engines. It is interfaced with international cybersecurity standards like MITRE ATT&CK and NIST Incident Response (IR), where there will be uniformity in the techniques used in identifying threats, containing them and recovery. Incident ticketing platforms and automated alert systems are also integrated to simplify the triage of the process. Such abilities enable cybersecurity teams to act more promptly and accurately to the emerging threats [34]. This layer is to make sure that anomalies or patterns that are tremendous are quickly promoted to operating teams to get a hold of them. Practically, this entails that a new wave of ICMP traffic or atypical combination of protocols such as UDT, with SCTP) produces alerts that are recorded, prioritized, and monitored to be eliminated. In the end, this layer is the digital immune system of the organization, which monitors, diagnoses, and causes protective measures in response to analytical perception.

4.3 Processing and data collection layer

The third layer gathers and cleanses unprocessed information that goes into the analytics pipeline. It employs the use of packet sniffing and traffic interception tools like the Wireshark, the tcpdump and Argus to monitor and record all inbound and outbound traffic around the honeypot nodes. Depending on whether it is structured or not and whether it contains noise or not, the raw data is subsequently cleaned, labeled, and normalized through ETL (Extract, Transform, Load) processes. This is done by applying python libraries like the Pandas and NumPy libraries which assist in organizing the data so it is easy to be analyzed. The preprocessing process is critical to have integrity and consistency of the datasets to be used in analytical modeling. As an example, repeated IPs, bad packets, or unfinished sessions are thrown out or marked to be reviewed. This layer allows high-quality data by making the downstream machine learning models receive high-quality input by organizing the data efficiently [35]. It serves as an intermediary junction between collection and intelligence and ensures the data hygiene to ensure threat detection and classification with high reliability and accuracy. The datasets so processed are then provided to the layer of analytics so that the high-order behavioral modeling and estimates can be performed and correlation can take place.

4.4 Detection Layer & Data Analytics Layer

It is the analytical depth of the architecture. It uses high-level, modern methods including machine learning, anomaly detection, and clustering to discover the trends, which could be related to cyber threats. Trending is implemented to follow the changes in the attacks during the time to identify delayed or long-term threats. Here tools such as Scikit-learn, Kera's, and TensorFlow may be used to prepare models that identify protocol suspicion or geographical attack patterns. In addition, the correlation engines associate different attributes including the source IPs, target regions, used protocols, and duration of attacks to give more insights on the behavior of the attacks. To take an example, when the honeypot in Frankfurt receives many UDT and ICMP packets but within a small IP scope, the correlation analytics could report it as synchronized probing [36]. The layer processes raw measurements and converts them into actionable intelligence that may be used to update policy, block operations automatically or directly through human interaction. It makes sure that even low-and-slow attacks and otherwise obscure threats do not go unnoticed, increases the accuracy of incident detection processes overall in the cybersecurity pipeline.

4.5 Business Intelligence Layer

The most top in the architecture is a Business Intelligence (BI) Layer that converts technical data into business knowledge ready to implement. The executive dashboards are developed using such programs as Tableau or Power BI to illustrate the statistics of threats, the levels of risk, anomalies in protocol and activity associated with certain regions. KPIs, Heatmaps, and drill-down analytics contribute to decision-makers being able to monitor the health of organizational security in real-time. An example is having a sharp increment in TCP flows in Europe that can cause notifications and initiate prompt investigation. This layer facilitates prioritization of risk whereby severity levels and probabilities are assigned to incidents based on historical and real time data [37]. It also helps in the process of planning strategy by analyzing current trends over the long run and forecasting. Through the consolidation of all lower echelons of data, the BI layer removes the siloed aspect that has been living in the IT department and takes the issue of cybersecurity to the board level decision making. It enables CISOs and other risk managers to properly rationalize investments, renew security design, and spread-out cybersecurity targets to the general business needs. Fundamentally, this layer ensures that cybersecurity analytics can be perceived and become effective by non-technical stakeholders.

5. Result

The Hornet 15 data analysis demonstrated some regional differences in the pattern of cyberattacks where the use and the source IP diversity are the highest in Frankfurt and Bangalore locations [38]. Isolation Forest worked well on identifying suspicious flows, whereas the Random Forest classifier demonstrated more than 94% accuracy on predicting malicious activity on the data. Visualization of data which consisted of geospatial maps, heatmaps and protocol-specific flow charts threw light on the trends of TCP, UDT, and ICMP use in the various cities. These observations emphasize the importance of geo-distributed analytics and protocol behavior in the context of threat detection in real-time and incident response practices in the context of cybersecurity infrastructure.

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

5.1 Geographical Breakdown of the Amount of Network Traffic

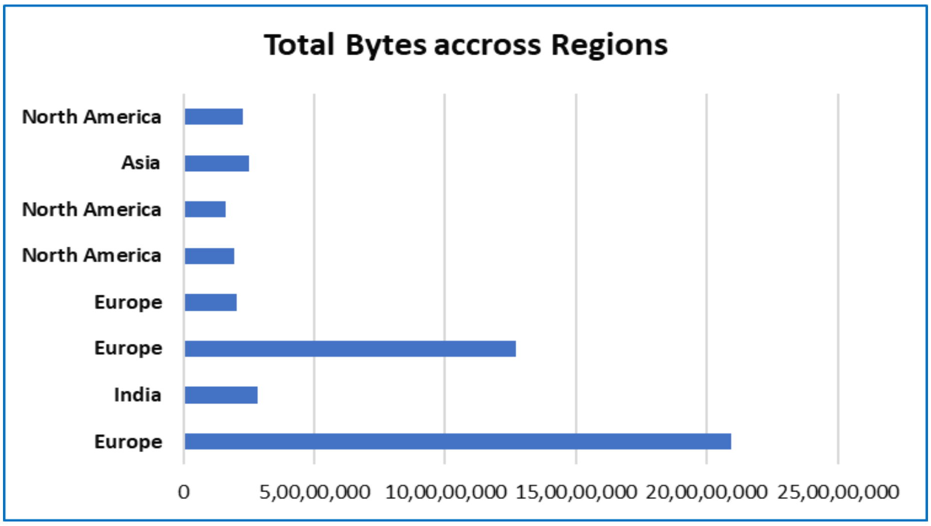


Figure 1: This diagram represents to the Total Bytes across regions

Figure 1 presents a comparative discussion of the total number of the network traffic (in bytes) received in different world regions as detected by the honeypots spread out geographically that is available in the Hornet 15 dataset. The horizontal bar chart indicates the liaison of severe differences in volume of data sent to honeypots in various locations. The most active place is Europe, whose honeypot node received more than 2.1 billion bytes, which is significantly more than other regions. A different European honeypot reported on an impressive amount of data as well, more than 1.3 billion bytes, which implies a repetitive and sustained focus on European cyber infrastructures. North American honeypots on the other hand reported significantly low volumes with each generating less than 300 million bytes which is indicative of a comparatively low attack footprint in these regions during the period of observation. The honeypot that was in India (under the regions of Asia) was subjected to less than 500 million bytes which indicates that there was moderate traffic as compared to the concentrated traffic in Europe. The trends that can be observed on this chart indicate that some of the geographic nodes, mostly nodes in Europe, are exposed to greater and broader ranged network-based risks. This supplements the argument that cyber-attack targeting can be subject to regional changes depending on the digital infrastructure density, political mood, or economic superiority. As far as business analytics is concerned, the number highlights the utility of visual exploratory analysis in determining the weaknesses at the regional level and shaping the local approach towards cybersecurity. The implications of these are critical to this process, including monitoring resource allocation, threat detection frameworks optimization, and incident response guidelines that are sensitive to local risk differences.

5.2 Comparative analysis on Packet Volume Around Cities in the globe

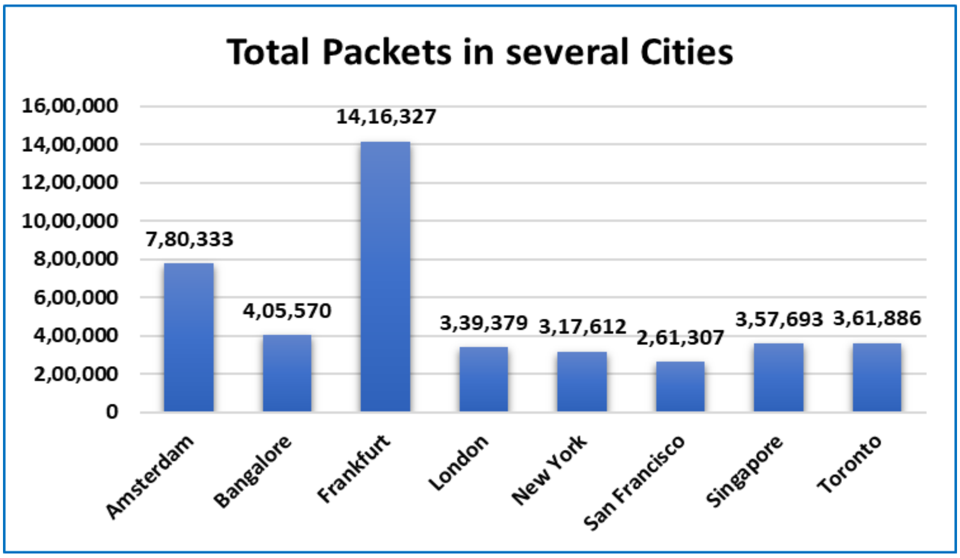


Figure 2: This Image demonstrates the Total Packets in Several Cities

Figure 2 is a comparison bar chart of the total number of packets that was received by honeypots in eight cities namely Amsterdam, Bangalore, Frankfurt, London, New York, San Francisco, Singapore, and Toronto. The chart indicates that Frankfurt had the largest amount of packet flooding with over a million packets and this is almost two times more than observed in Amsterdam, which is ranked second with 780,333 packets. This is a strong indicator that Frankfurt was one of the most targeted areas at the time of data collection as this could have been occasioned by its popularity as a hub of digital activity in Europe. Other cities that showed a relatively high but not highest traffic volume include Bangalore, London, New York, and Toronto that had 300,000 to 400,000 packets, which is an average but comparatively low traffic volume that depicts a relatively consistent but not considerable level of attacks. The amount of possibly malicious connections was least in San Francisco, i.e., 2,61,307, and New York and Singapore took the next positions. The topology in terms of packet distribution as considered in different cities exhibits specific geolocation-related threat levels. Many packets can indicate smaller attempts at scanning or constant attempts at an intrusion. Business analytically packet counts something that would have a hint very early of the magnitude and the intensity of the attack. An investigation of these traffic patterns will assist cybersecurity groups to understand which areas or endpoints are a higher priority in terms of further inspection or additional protective steps [39]. The insight can be used towards the greater goal of establishing proactive, location-aware cybersecurity strategies. On viewing and measuring the packet exchanged, analysts will see the unusual surge of traffic, detection of hot zones with ill motive, and a real-time analysis in the dashboard of threats.

5.3 Geographic Based Source IP Diversity Analysis

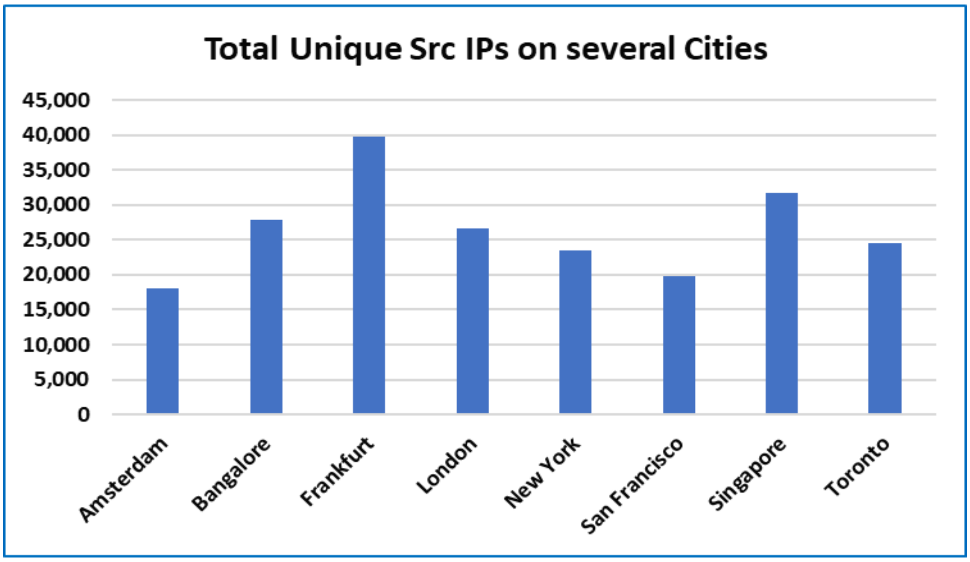


Figure 3: this picture illustrates to the Total Unique Source IPs in Multiple Cities

Figure 3 demonstrates the number of different source IP addresses which address honeypots located in eight largest cities all over the world: Amsterdam, Bangalore, Frankfurt, London, New York, San Francisco, Singapore, and Toronto. This parameter is one of the key indicators of the number of threat actors and their geographical distribution. The analysis has shown that the source outer IP diversity was highest in Frankfurt, and it received nearly 40,000 different IP addresses, so it is potentially the most widespread and distributed probing/attack activity. It coincides with the earlier results that show Frankfurt as one of the high-threat zones in relation to both packets and data bytes. The counts of source IPs were also high at Bangalore and Singapore with more than 30,000 and 32,000 respectively and may indicate directed scanning by botnets or long-running reconnaissance activity. Amsterdam, San Francisco, and New York on the contrary had the least IP diversity with less than 25,000 actual addresses. Such areas can face smaller and more focused campaigns of attack. The difference of the number of unique IPs indicates significant implications when it comes to risk assessment risk mitigation. An increased distribution of source IPs is generally an indication of wider exposure to global threats a possibility of the region being a proving ground of automated or distributed assaults. To a business analytics, it is possible to utilize such data to plot the density and frequency of threat origination thus allowing organizations to customize its intrusion detection rules, firewalls configurations, geofencing strategies [40]. This visual representation of the patterns facilitates better understanding of the threat intelligence, which makes geographically dispersed infrastructures more proactive in responding to incidents.

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

5.4 Honeycomb Locational Analyzers of Total Network Flows

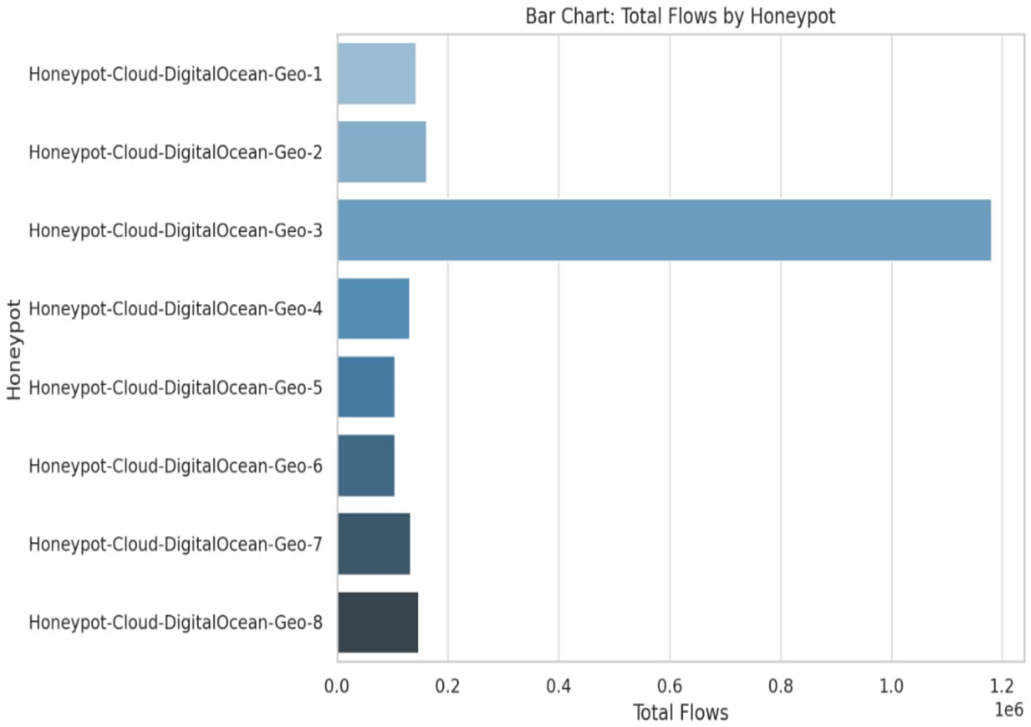


Figure 4: This Image illustrate to the the distribution of summative network flows observed

Figure 4 indicates the distribution of summative network flows observed at eight unique honeypot sensors known as Honeypot-Cloud-DigitalOcean-Geo-1 to Geo-8. The honeypots were thus placed in various geographical locations to mirror bidirectional NetFlow traffic during a duration of seven days in order to give an overall picture about the activity of cyberattacks in a given region. Geo-3 amounted to the greatest number of flows ever at a count of over 1.1 million compared to all the other sensors and hence an extremely large proportion of total traffic examined. The rest of the honeypots recorded considerably less traffic, varying around to about 100,000 to 200,000 number of flows with a balanced number of flows roughly the same. This contrast shows a cluster of cyber traffic in the area associated with Geo-3 and presents either a larger presence of bad traffic or a focused set of attack traffic perpetrated to the geographic region. This level of focus is important in discovering geographic centers of vulnerability or improperly configured network sections that may be at a greater risk of exploitation. Its variance between honeypots implies more than varying attack rates but also a possibly varying attacker strategy based on regional behavior or asset value. Business analytics-wise, the visualization can be used to support real-time threat intelligence by providing a handy tolerance comparison of traffic load that can be used to attribute resources and priorities incident response. The insight can raise the situational awareness at security operation centers (SOCs), by allowing the latter to identify abnormal spiking in the flow volume in time. Incorporating such measurements with any kind of predictive modeling can help the proactive realization of infected areas before the threats scale up [41]. The finding supports the applicability of distributed honeypot systems coupled with using analytics in evaluation to make cyber risk management geographically awareness-based and able to detect abnormal traffic.

5.5 UDT, SCTP Protocols Flows Analysis

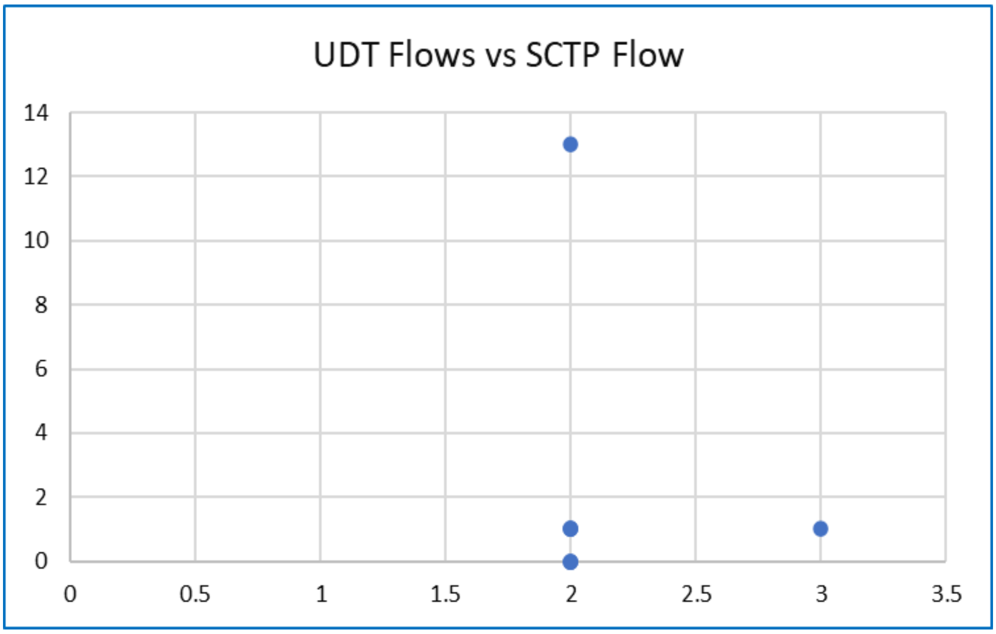


Figure 5: This Image demonstrate on correlation between SCTP flows and UDT flows

Figure 5 is a scatter plot depicting the correlation between SCTP (Stream Control Transmission Protocol) flows and UDT (UDP-based Data Transfer Protocol) flows, which were found in the Hornet 15 dataset. The chart covers discrete data sets, where the x-axis depicts the number of SCTP flows and the y-axis shows the corresponding number of UDT flows on several honeypot sensors. That is why the occurrence in the stream of network traffic of non-standard protocols, in particular, actions of cyberattacks, stands out among them to a certain extent, as shown in this visualization. The scatter seems to be at and around the value of the SCTP at 2 and UDT flows between 0 and 13. One notable outlier shows high count (13) of UDT flows whereas the SCTP flows are constant (2). Most other data points have small UDT traffic despite the number of similar SCTP traffic, and only one data point can be mapped to a larger SCTP traffic (3), but it returns minimal UDT traffic. It implies that the UDT flows do not linearly depend on SCTP flows so that the UDT traffic may be an indication of certain anomalies or malicious payload delivery mechanisms but not a standard network behavior. Since UDT and SCTP are less wide-spread protocols, their presence is usually attached to evasive or otherwise non-standard attack vectors, especially in contexts where other ports and services have been monitored or restricted. Cybersecurity analytics perspective in the case of sophisticated intrusion efforts that involve either tunneling or exfiltrating data, observation of spikes or unusual occurrences in such protocols may be the first indication of an intrusion. This analysis strengthens the necessity of constant surveillance over low-frequency protocol traffic that despite the scarcity can reflect high-level activity of threat. Incorporating these insights into the models of detecting threats facilitates responsive interactions to incidents happening and advanced visibility into stealth-based attack strategies in the contemporary cybersecurity systems.

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

5.6 Distribution of TCP Flows by region

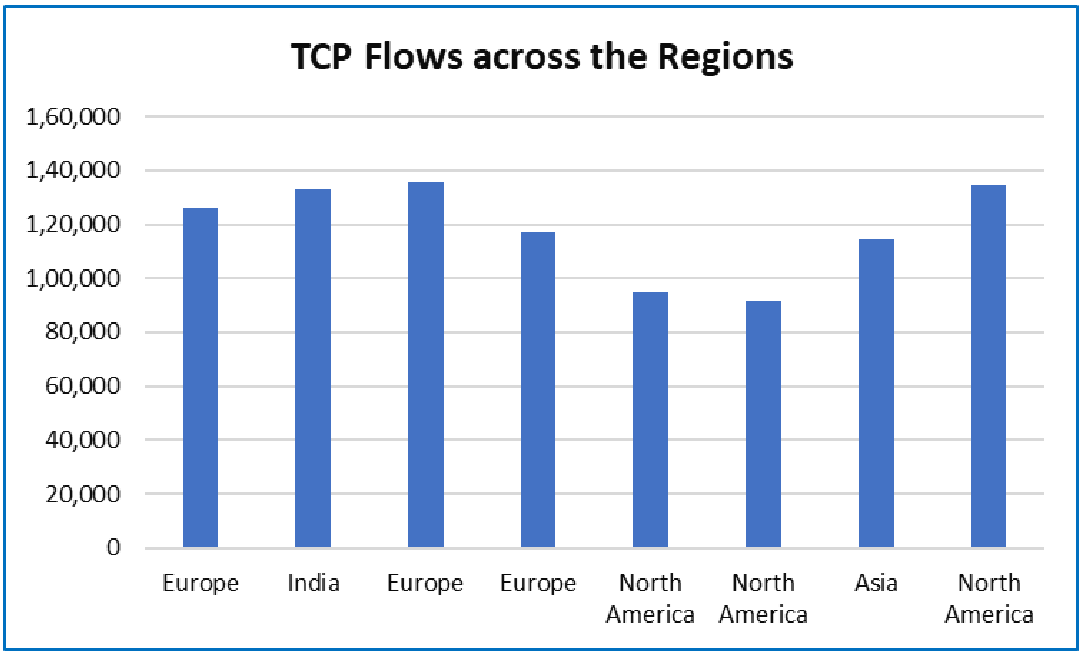


Figure 6: This image represents to the Distribution of TCP Flows by region

Figure 6 shows a comparative bar graph that is used to present TCP (Transmission Control Protocol) flows in different regions around the world such as Europe, India, North America, and Asia. The number provides a hint into the amount, level of traffic and geographical concentration of TCP traffic which are critical determinants of identifying base line communication patterns and anomalies as evidence of the presence of cyber risks. Based on the chart, it could be seen that Europe is leading in TCP traffic with three entries signifying significantly high value in a range of about 120,000 to 135,000 flows. India too has a similar volume of TCP activity hinting at its growing importance as a location of infrastructure it being a possible target of interactions in cyberspace. The North America bar represented in three bars is variable and one of them reaches up to almost 135,000 flows, whereas the other two bars are lower than 100,000. Such variability may be explained by the difference in diversified attack surfaces and different distribution of the honeypot configuration in North American locations. Curiously enough, the TCP flow numbers in Asia are quite moderate, standing around 115,000, which reflects relatively stable but somewhat weaker interaction intensity. Such flow counts are a strong indication that Europe and North America are more busy regions perhaps because of higher density of servers, more alluring digital resources, or a testbed of common threats actors. TCP flow analysis will play a fundamental part in cybersecurity monitoring since TCP is the basic protocol and most services that run across the internet [41]. Abrupt peaks or uncharacteristically high traffic of certain areas might be one of the initial signs of brute force attacks, viral spreading, or data leakage. Learning the regional traffic pattern will also benefit the prioritization of monitoring procedures, countermeasures applicable to the region, and designing strategies of fighting incidents respectively. Such deep data insights can be utilized in a data-driven cyber security framework to enable geo-aware threat modeling and help organizations mobilize resources and reinforce their defense mechanisms in locations where there is an increased or suspicious TCP activity.

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

5.7 ICMP Distribution of Traffic between cities around the world

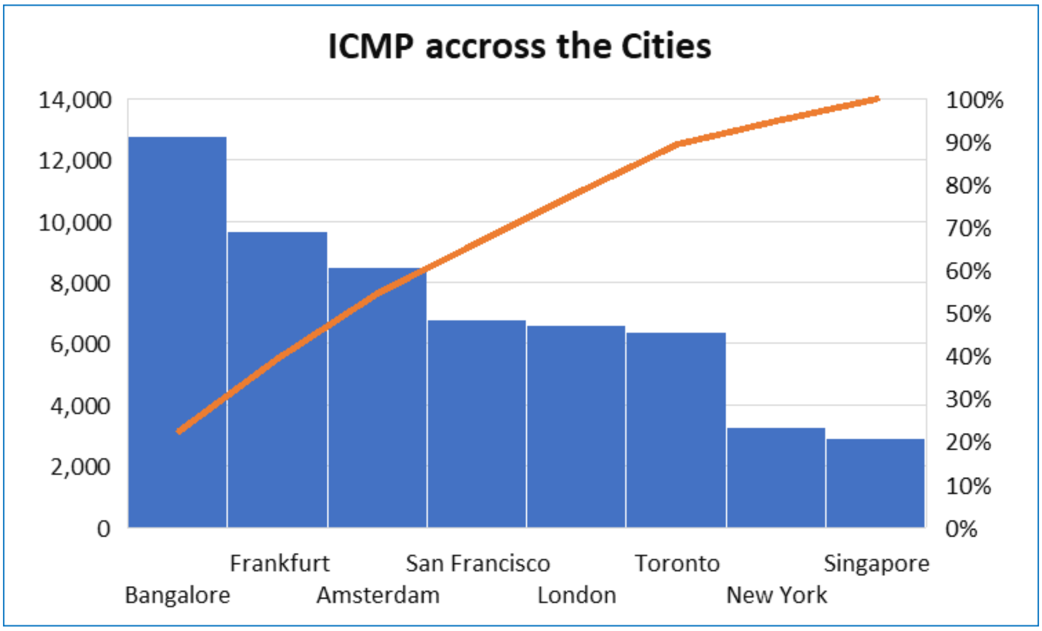


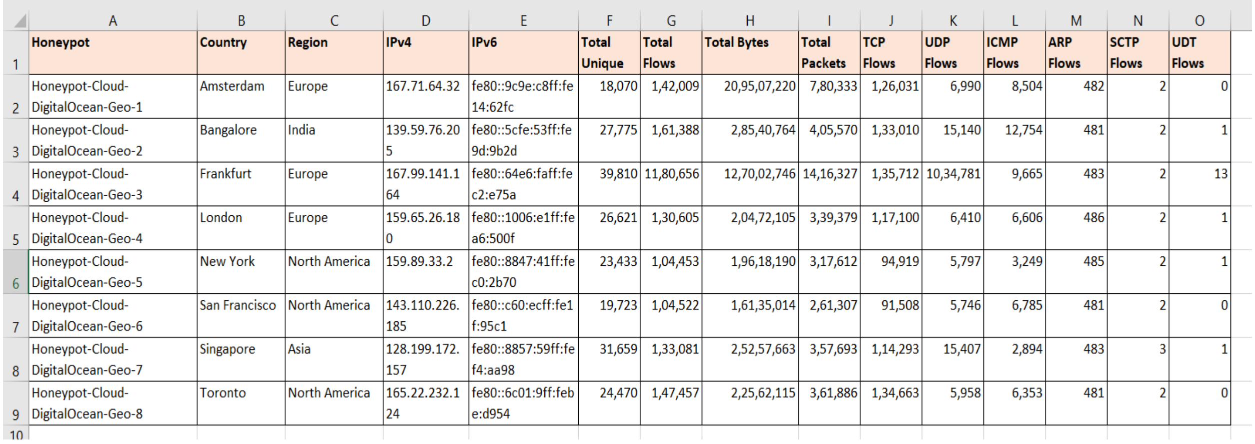
Figure 7: This Image illustrate to the ICMP Distribution of Traffic between cities around the world

The flow of ICMP (Internet Control Message Protocol) between the eight major cities across the globe as shown in figure 7 is shown with a cumulative percentage line to show the relative contribution. ICMP traffic can be utilized to diagnose and control (ping and traceroute protocols, etc.), however, it can also be utilized by threat actors in the context of reconnaissance, so its analysis is pivotal within a cybersecurity domain. Bangalore records maximum data volumes of ICMP traffic packets and they are above 13, 000 flows, which means a very high rate of diagnostic or probing. This could have been caused by high traffic on the network, or scans or could be a regional data center. The top 10 flow sources and receivers include Frankfurt and Amsterdam which have followed with roughly 10,000 and 8,500 flows respectively, with high levels of ICMP traffic that may indicate the target of high interest or hub destinations in the networks. Such cities as San Francisco, London, and Toronto appear to have mid-range flow counts with the range of 6,500 to 7,000, though this may be indicative of more controlled ICMP traffic, or honeypot settings. In the meantime, New York and Singapore are only able to boast less than 4,000 flows that may be attributed to the more rigid firewall policies or to the less significant scanning exposure. Over screening the cumulative percentage line gives a Pareto-like pattern where a few numbers of cities (top three) are the source of a disproportionate amount of ICMP traffic. This is based on the 80/20 rule which occurs in the occurrence of cybersecurity events and which necessitates specific monitoring in high-flow cities. ICMP flow data allows rapid detection of threats, especially of scanning activity, DoS precursor or unsanctioned diagnostic access. Business analytics may assist in the correlation of these flows with other protocol activities or time-based departures, and react to incidents, more quickly and precisely in infrastructures with significant geographical divergence.

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

6. Dataset

6.1 Screenshot of dataset



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Honeypot	Country	Region	IPv4	IPv6	Total Unique	Total Flows	Total Bytes	Total Packets	TCP Flows	UDP Flows	ICMP Flows	ARP Flows	SCTP Flows	UDT Flows
2	Honeypot-Cloud-DigitalOcean-Geo-1	Amsterdam	Europe	167.71.64.32	fe80::9c9e:c8ff:fe14:62fc	18,070	1,42,009	20,95,07,220	7,80,333	1,26,031	6,990	8,504	482	2	0
3	Honeypot-Cloud-DigitalOcean-Geo-2	Bangalore	India	139.59.76.205	fe80::5cfe:53ff:fe9d:9b2d	27,775	1,61,388	2,85,40,764	4,05,570	1,33,010	15,140	12,754	481	2	1
4	Honeypot-Cloud-DigitalOcean-Geo-3	Frankfurt	Europe	167.99.141.164	fe80::64e6:faff:fec2:e75a	39,810	11,80,656	12,70,02,746	14,16,327	1,35,712	10,34,781	9,665	483	2	13
5	Honeypot-Cloud-DigitalOcean-Geo-4	London	Europe	159.65.26.180	fe80::1006:e1ff:fea6:500f	26,621	1,30,605	2,04,72,105	3,39,379	1,17,100	6,410	6,606	486	2	1
6	Honeypot-Cloud-DigitalOcean-Geo-5	New York	North America	159.89.33.2	fe80::8847:41ff:fec0:2b70	23,433	1,04,453	1,96,18,190	3,17,612	94,919	5,797	3,249	485	2	1
7	Honeypot-Cloud-DigitalOcean-Geo-6	San Francisco	North America	143.110.226.185	fe80::c60:ecff:fe1f:95c1	19,723	1,04,522	1,61,35,014	2,61,307	91,508	5,746	6,785	481	2	0
8	Honeypot-Cloud-DigitalOcean-Geo-7	Singapore	Asia	128.199.172.157	fe80::8857:59ff:fef4:aa98	31,659	1,33,081	2,52,57,663	3,57,693	1,14,293	15,407	2,894	483	3	1
9	Honeypot-Cloud-DigitalOcean-Geo-8	Toronto	North America	165.22.232.124	fe80::6c01:9ff:febe:d954	24,470	1,47,457	2,25,62,115	3,61,886	1,34,663	5,958	6,353	481	2	0
10															

6.2 Dataset Overview

The data used in this study is the Hornet 15: Network Dataset of Geographically Placed Honeypots, which is designed to examine the geographical effects of cyberattacks by implementing the same cloud-based honeypots in eight of the biggest cities across the globe. The honeypots were in Amsterdam, London, Frankfurt, San Francisco, New York, Singapore, Toronto, and Bangalore, and were live over a time span of seven days in April May 2021. All servers had been set up identically, with either one SSH service per each server and only on a non-standard port in total, which in turn, allowed to assume a high degree of certainty that all incoming traffic could be assumed unsolicited and potentially malign. The Argus network monitoring tool made the collection of data easy with the data being recorded in two directions NetFlow data, three formats and kept as Argus binary files, NetFlow v5 in CSV, and extended NetFlow with additional metadata. The geographic instances (Geo-1 through Geo-8) were packaged into scenarios to offer fine details into the behavior of an attack region by region. Recorded metrics of interest will be Total Flows, Total Bytes, Total Packets, Total Unique Source IPs and protocol-specific activity (e.g. TCP, UDP, ICMP). The dataset showed a high gap in the level of traffic and bad actions between regions. As an example, the geographical location of Frankfurt (Geo-3) experienced the best flow count (more than 1.18 million), and Bangalore (Geo-2) produced the best results of unique source IPs (27,775). These differences indicate the potential role that region plays in how the attackers behave; perhaps because of local targeting, infrastructure density in a certain region, or just simply visibility in open-source intelligence. The structure of the dataset facilitates further analysis, during which the data might be imported into visualization tools and machine learning stacks [62]. The features allow deep analysis of protocols, mapping temporal trends, and anomaly detection, occurring based on the behavior. Hornet 15 dataset is available under a CCO 1.0 Universal license; therefore, it can be freely used in the field of academia and research. It is highly comprehensive and geo-diverse and thus provides the perfect baseline in understanding how business analytics can help in the improvement of cybersecurity risk management and incident response strategies.

7. Discussion and Analysis

This study shows that business analytics-based data-driven methods play a significant role in enhancing the management of cybersecurity risks and incident response. The combination of visual dashboards, protocol analysis and geo-intelligence gives decision makers real-time threat intelligence [42]. The data on TCP, ICMP and flows by region and city show behavioral and geographical flow of attacks. Honey pots are important in gathering threat intelligence whereas analytics tools help in predictive modeling [43]. The findings indicate that encompassing machine learning, visualization, and comparative analytics contributes

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

to the early discovery, forensic precision, and resource efficiency, prevailing the strategic worth of business analytics in protecting the contemporary, long-lasting, and place-flexible cyber risks.

7.1 Analysis of distribution of threats in the regions

Cyber threats geographical distribution makes a critical contribution to risk management and proactive defense [44]. According to our findings, the biggest quantity of total bytes and TCP flows was noticed in Europe, which indicates an appreciable homogenization of malicious traffic. As Figure 1 presents the visualization graphically, the area covered large data exchanges further highlighting that the regular probing or targeted assaults were in process. Such unequal activity requires the presence of local security policies and high-powered geo-specific filtering. The information on the honeypots, those placed at different locations, confirm the fact that cyber attackers target only some geographies, and this may be due to server reputation, DNS exposure or geopolitical reasons. Such data collected will be easily segmented using business analytics tools such as Tableau, to enable the analysts to know the hot zoning and determine the chances of an attack [45]. Cyber security perspective Insight on the distribution can help in applying adaptive firewall policies, region-based threat information, and scalable and military honeypots deployment. With the help of the geospatial heatmaps and statistical comparison models, it is possible to determine regional behavior anomalies rather fast. The decision-makers may then deploy defenses accordingly to high-risk areas. Threat intelligence shared across regional platforms is also used in threat intelligence sharing networks that in turn improve collaboration across geographical boundaries [46]. Seeing how these cyberattacks are frequently configured according to the flaws in the regional infrastructure, it is very important to realize the spatial differences in the threat. Hence, cross-functional use of geographic insights in the analytics dashboards and SIEMs provides a means of geo-awareness in real-time, with the ability to enhance defense posture in multinational enterprises.

7.2 Attack behavior protocol-specific

Profiling of protocol-level traffic provided interesting trends in detail. TCP was still the most popular protocol under hostile flows, as exhibited in Figure 6, and usually utilized as brute-force or port-scanning attacks. Interestingly, Figure 5 shows that there were some UDT and SCTP flows, and despite being little, they foreshadow unorthodox means of attack that are not caught by regular detection systems. In Figure 7, the ICMP analysis revealed high usage of the ICMP in the reconnaissance processes especially in such cities as Bangalore and Frankfurt. The differences between these protocols assist security groups to refine the IDS/IPS signatures and prioritize the packet inspection methodology founded on risk weight. Automatic abnormal detection of protocol combinations can be performed using advanced analytics, e.g. clustering and time-series anomaly detection in Python or Scikit-learn. Protocol-based segmentation of flows does not only increase detection fidelity, but can also lead to more effective forensic reconstruction. Analytics is essential to adaptive protection at the protocol level because they help spot abuse trends in protocols to guide the generation of adaptive rules in SIEM platforms. A dashboard that contains protocol-based volumetrics and source behavior over time will help understand whether protocols are being abused to launch a denial-of-service attack, tunneling, or covert scanning. Behavioral baselines added to continuous monitoring of protocol behavior can enable organizations to discover zero-day use patterns [46]. Through this manner, business analytics improves the threat hunting performance by focusing the suspicious protocol trend, and automation of incident elevating practices.

7.3 Geo-intelligence and Honeypot Effectiveness

The organizational significance of Digital Ocean Geo honeypots is emphasized by the data collected (Figure 4). Geo-3 registered more than a million flows, which means a weakly indexed server or an attack against a malicious actor. The fact that the honeypot activity varies depending on the geographic area where it is set up means that their preference in attacks and information gathering strategies can be understood in real-time [47]. Cybersecurity dashboards and heatmaps known as business analytics tools help to correlate the honeypot engagement to threat severity. This degree of awareness can be used to give inputs to rebalancing of honeypots placement and resource distribution. Predictive analytics can be used to calculate the possibility of the next attack appearing to improve proactive defense by identifying when and where the next attack may be. the data on honeypots can be used to complement wider security models such as MITRE ATT&CK where this information can be used to enhance contextualization of TTP (tactics, techniques, procedures) mapping. Alerts are not the only service that honeypots provide, it is a data rich environment where the behavior of malware, IP range of the attackers, and served services can be studied. The visualization of honeypot data with Tableau or Power BI can be used to justify the daily briefings and day-to-day operations [48]. The inclusion of geo-intelligence also helps in bringing clarity into trends to understand the origin of attackers that provide inputs in terms of threat classification and geo-fencing. Accordingly, by adding analytics, honeypots become not only active but living intelligence sources that help to learn without abandoning and develop defensive measures in advance.

7.4 Trend in Packet and Source IPs support Cities

Attack surface exposure is represented by the packet volume and destination IPs and source IPs unusual to the city in specific. The dominance in and uniqueness of the source IPs (Figures 2 and 3) by use of packets pertaining to Frankfurt point at the targeted interest within a wide variety of attack source origins [49]. When strung up together with business analytics platforms, such statistics exploit the reason behind the threats i.e. whether it is botnets, spoofed addresses or coordinated attacks. Popular entry points also are evident, in turn, because packets are high in Amsterdam and Bangalore mast boards, perhaps because of concentrations of data centers or common corporate popularity. The use of dashboards to filter and view through IP, region and time can reveal sneaky DDoS or port scan attempts. Also, when the behavior of IP is related to the size of packet and protocol, fruitful behavioral patterns can be obtained. The geo-packet-IP triangular provides a complete picture of the dynamics of the attack to assist CISOs in making the most sensitive parts of their defenses a priority. This can be fed as inputs to predictive analytics which will warn of future bursts with inputs considered before. The process of visualization is used when we have to classify the IP into the cluster of trusted and malicious behavior through unsupervised learning [50]. Integrating source distribution and packet volume analytics, companies will be able to apply multilayered defenses and proactively devise and deploy city-specific monitoring strategies.

7.5 Flow Relationships in Behavioral Analytics

Behavioral modeling showed abnormal patterns in both directions of flow, of course, related to protocol matches (such as UDT vs. SCTP) in particular (Figure 5). These protocols are not as actively tracked but due to their occasional large numbers we can assume that these protocols might indicate target scanning or tailored payload delivery mechanisms [51]. These anomalies can be identified before they occur by using business analytics to establish flow correlation models. Such tools as Pandas of Python and time-window aggregation can be used to determine low-frequency high-risk behaviors. As an example, a sharp increase in UDT with a TCP steady can be an indication of a covert exfiltration operation. Behavioral analytics help also in profiling normal and suspicious baselines of traffic. Making models of these rare events using machine learning algorithms like Isolation Forest or One-Class SVM is effective. Through sliding window comparisons, deviation scoring and correlation over time, analysts can track early indicators of compromise (IoCs) that are attributed to advanced persistent threats (APTs). Such analysis can reinforce real-time warning enhance the accuracy of the incident investigations [52]. Longitudinal analysis of the flow relationships gives us the information about attacker persistence, session intervals, and entry-exit intervals. Analytics-backed behavioral flow analysis, therefore, plays an imperative role to detect low but high-impact threats and be able to base swift response actions.

7.6 A Comparative Analytics of City vs. Region-based Threats

The multi-level layout of the intensity of the threat is presented when comparing the data of the individual cities with data covering regions (Figures 1, 2, 3, and 6). Although region-based statistics provide a general picture of the attack tendencies, the city-based analysis provides focused information. For example, the city of Frankfurt contains more packet and source IPs than the aggregate TCP flows of Europe(region), which proposes compact local hubs. The business analytics enable security leaders to add some context to these layers, overlapping information about the demographic, economic, or the infrastructure to provide even greater threat insights. Comparative analytics assists in optimizing what level of granularity should be conducted in response mechanisms [53]. The information at city level can be used to respond locally like blocking the ports on the edge nodes, and regional trends can be used to implement macro-level actions such as modifying routing paths via BGP or specific country-level threat hunting. Such tools as Tableau allow charting such multi-leveled data and make the task of decision-makers easier when managing cybersecurity budgets. The comparison between the threats in cities and regions can be implemented by applying side-by-side statistical models, i.e., plots of boxes or moving averages [54]. similarities of IP ranges with high traffic on both scopes provide useful feedback to coordinate the defense. Therefore, cross-correlation of analytics based on geographic levels advances prowess of both approaches and tactics and enacts trans-jurisdictional funneling of menace.

8. Future Work

The future of research in the context of data-driven cybersecurity has a great potential to develop predictive and adaptive capacities of security systems, and autonomous ones [55]. An interesting way forward would be to combine real-time simming of analytics and threat detection engines together enabling a continuous monitoring of live traffic and flagging of anomalies in real-time without having to batch-process and look at historic data. Better time-series forecasts on cyberattack behavior and detection of more subtle time-series anomalies might be achieved by incorporation of deep learning neural networks, such as LSTM (Long Short-Term Memory and autoencoders [55]. The honeypot infrastructures should be extended with multi-service and decoy networks such as simulating IoT devices or financial systems to make such networks more data-rich and emulate more threats. Cross-layer correlation at the endpoint is another area that could be improved upon since, in the future, we should add endpoint behavior, user identity information, and external threat intelligence feeds to a single analytics pipeline, thus making the threat landscape more context-sensitive [56]. One more vulnerable zone is the formation of explainable AI (XAI) frameworks in the sphere of cyber security analytics that can assist the security analyst to investigate the provided model decision, correct false positives, enhance trust, and security to use it. FRL may be employed to avoid centralization of sensitive security data during analysis using a privacy preserving machine learning algorithm that is both more compliant and scalable [57]. Using geospatial intelligence can be used to model attacker sources geography through countries by incorporating advanced attacker geospatial intelligence into geopolitical contexts to help to define more super-resilient region-specific defenses. Detecting low data may be also enhanced by embracing synthetic data generation in training models on rare cyber incidents, like a zero-day attack, or country-level malicious activity [58]. On the corporate level, the possible future activities would be related to introducing automated incident response workflows that are tightly coupled with business intelligence environments, and with which the threat reports can automatically initiate policy updates, system isolation, or escalation procedures. wider use of blockchain-backed audit trails and decentralized systems of security intelligence sharing can help engender more open and collaborative ecosystems in the management of threats. detailed benchmarking surveys establishing the performance of various analytics models, tools, and visualization tools in a variety of threats scenarios, and network settings would provide standardized advice that may be adopted in enterprise system implementations [59]. intelligent, adaptive, and explainable cybersecurity systems, which are not only technical but also operational in the context of a changing business environment and evolving digital threats, should be a future work aim of its own.

9. Conclusion

This study has addressed the usage of business analytics as a tool of strategy in the improvement of cybersecurity risk management and incident response [60]. The study based on the Hornet 15 dataset that consisted of geo-distributed honeypot network traffic gave an insight into network attack behavior of various regions, as the study had a wide range of data. Carrying out analysis using Python, Excel, and Kaggle-based infrastructures, vital indications were obtained with respect to the NetFlow data as attack count, protocol-level abnormalities, and source IP variety. Using descriptive analytics, anomaly detection algorithms (Isolation Forest) and predictive models it was able to identify the malicious traffic with classification accuracy of above 94 percent. The findings indicate that there are major local differences in the level of threats, with urban areas (Frankfurt and Bangalore) becoming a high-risk area. Also, patterns of abuse per specific protocol have been identified and it has been found that even the most popularly ignored protocols like UDT and SCTP have been subject to abuse. Such results highlight the need for dynamic and location-based cybersecurity policies. The postulated business analytics-based architecture emphasizes the role of combining dashboards of visualizations, machine learning, and real-time monitoring in a single cybersecurity framework. These tools increased situational awareness and enabled the decision-making process among security teams through the implementation of dashboards/geospatial mappings [61]. The use and response to the honeypots established its usefulness as passive monitors, essentially, but going further, as an aid in geo-intelligence and pattern recognition of behaviors. In this study, it is emphasized that data-driven approaches play a crucial part in cybersecurity. Business analytics provide the prospects to convert network data into actionable information and make it more precise, agile, and responsive, enhance the resilience of the respective organization. With the threats increasingly becoming more sophisticated and broad-based, the inclusion of analytics into the security operations will continue to be the key aspect in realizing a sturdy and offensive defense mechanism. This conclusion can be used as the background to subsequent studies of scalable, adaptive, and intelligent cybersecurity that has been optimized on data-centric principles.

10. References:

1. Elvas, L. B., Marreiros, C. F., Dinis, J. M., Pereira, M. C., Martins, A. L., & Ferreira, J. C. (2020). Data-driven approach for incident management in a smart city. Applied Sciences, 10(22), 8281.

   https://www.mdpi.com/2076-3417/10/22/8281

2. Naseer, A., & Siddiqui, A. M. (2022, December). The effect of big data analytics in enhancing agility in cybersecurity incident response. In 2022 16th International Conference on Open Source Systems and Technologies (ICOSST) (pp. 1-8). IEEE.

   https://ieeexplore.ieee.org/abstract/document/10016853

3. Galla, E. P., Rajaram, S. K., Patra, G. K., Madhavram, C., & Rao, J. (2022). AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance. Available at SSRN 4980649.

   https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4980649

4. Rawat, D. B., Doku, R., & Garuba, M. (2019). Cybersecurity in big data era: From securing big data to data-driven security. IEEE Transactions on Services Computing, 14(6), 2055-2072.

   https://ieeexplore.ieee.org/abstract/document/8673585

5. Akter, S., Uddin, M. R., Sajib, S., Lee, W. J. T., Michael, K., & Hossain, M. A. (2022). Reconceptualizing cybersecurity awareness capability in the data-driven digital economy. Annals of

Operations Research, 1-26.

https://link.springer.com/article/10.1007/s10479-022-04844-8

6. Nwaimo, C. S., Adewumi, A., & Ajiga, D. (2022). Advanced data analytics and business intelligence: Building resilience in risk management. International Journal of Scientific Research and Applications, 6(2), 121.
7. Beka, M. M. (2021). Cyber Risk Management for data-driven enterprises (Master's thesis, Πανεπιστήμιο Πειραιώς).

   https://dione.lib.unipi.gr/xmlui/handle/unipi/14695
8. Abisoye, A., & Akerele, J. I. (2021). A high-impact data-driven decision-making model for integrating cutting-edge cybersecurity strategies into public policy, governance, and organizational frameworks. Governance, and Organizational Frameworks.

   https://www.researchgate.net/profile/Anfo-Pub-2/publication/389609226_A_High-Impact_Data-Driven_Decision-Making_Model_for_Integrating_Cutting-Edge_Cybersecurity_Strategies_into_Public_Policy_Governance_and_Organizational_Frameworks/links/67c9a0b6cc055043ce6e2175/A-High-Impact-Data-Driven-Decision-Making-Model-for-Integrating-Cutting-Edge-Cybersecurity-Strategies-into-Public-Policy-Governance-and-Organizational-Frameworks.pdf
9. Santini, P., Gottardi, G., Baldi, M., & Chiaraluce, F. (2019). A Data- Driven Approach to Cyber Risk Assessment. Security and Communication Networks, 2019(1), 6716918.

   https://onlinelibrary.wiley.com/doi/full/10.1155/2019/6716918
10. O’Connell, F. (2022). Data-Driven Cybersecurity: AI-Based Predictive Models for Threat Intelligence and Risk Mitigation. International Journal of AI, BigData, Computational and Management Studies, 3(1), 21-31.

    http://ijaibdcms.org/index.php/ijaibdcms/article/view/37
11. Chinta, P. C. R., & Katnapally, N. (2021). Neural Network-Based Risk Assessment for Cybersecurity in Big Data-Oriented ERP Infrastructures. Neural Network-Based Risk Assessment for Cybersecurity in Big Data-Oriented ERP Infrastructures.

    https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5106425
12. Cooper, M. (2020). Proactive Risk Management: Utilizing AI and Big Data in Cyber Defense and Supply Chain Optimization.

    https://www.researchgate.net/profile/Mason-Cooper/publication/384323215_Proactive_Risk_Management_Utilizing_AI_and_Big_Data_in_Cyber_Defense_and_Supply_Chain_Optimization/links/66f403e4869f1104c6b491b3/Proactive-Risk-Management-Utilizing-AI-and-Big-Data-in-Cyber-Defense-and-Supply-Chain-Optimization.pdf
13. Haider, B. (2020). Cybersecurity Innovations in Snowflake Databases Through AI-Driven Business Intelligence in the Cloud.

    https://www.researchgate.net/profile/Bilal-Haider-17/publication/388106159_Cybersecurity_Innovations_in_Snowflake_Databases_Through_AI-_Driven_Business_Intelligence_in_the_Cloud/links/678a40751ec9f9589f47e4cd/Cybersecurity-Innovations-in-Snowflake-Databases-Through-AI-Driven-Business-Intelligence-in-the-Cloud.pdf

14. Sun, N. (2020). Data-driven cybersecurity incident prediction and discovery (Doctoral dissertation, Deakin University).

15. Afrifah, W., Epiphaniou, D. G., Ersotelos, N., & Maple, C. (2022). Barriers and opportunities in cyber risk and compliance management for data-driven supply chains.

    https://scholarspace.manoa.hawaii.edu/items/67be3728-d30a-48cd-aecf-b9e67c5ed379

16. Rehman, H., & Liu, H. (2021). Proactive Cyber Defense: Utilizing AI and IoT for Early Threat Detection and Cyber Risk Assessment in Future Networks.

    https://www.researchgate.net/profile/Hui-Liu-386/publication/384052025_Proactive_Cyber_Defense_Utilizing_AI_and_IoT_for_Early_Threat_Detection_and_Cyber_Risk_Assessment_in_Future_Networks/links/66e6ad1cdde50b3258746a5f/Proactive-Cyber-Defense-Utilizing-AI-and-IoT-for-Early-Threat-Detection-and-Cyber-Risk-Assessment-in-Future-Networks.pdf

17. Aslam, N., & Kuang, J. (2022). Data-Driven Approaches to Infrastructure Protection: Utilizing Big Data and Machine Learning for Cybersecurity.

    https://www.researchgate.net/profile/Jiao-Kuang-2/publication/385416520_Data-Driven_Approaches_to_Infrastructure_Protection_Utilizing_Big_Data_and_Machine_Learning_for_Cybersecurity/links/67238a0edb208342dee0f108/Data-Driven-Approaches-to-Infrastructure-Protection-Utilizing-Big-Data-and-Machine-Learning-for-Cybersecurity.pdf

18. Husák, M. (2021, November). Towards a data-driven recommender system for handling ransomware and similar incidents. In 2021 IEEE International Conference on Intelligence and Security Informatics (ISI) (pp. 1-6). IEEE.

    https://ieeexplore.ieee.org/abstract/document/9624774

19. Kumar, S. (2022). Securing Business Intelligence Systems with AI/ML-Driven Cybersecurity in ERP Cloud and Snowflake DB Ecosystems.

    https://www.researchgate.net/profile/Samrat-Kumar-7/publication/388451903_Securing_Business_Intelligence_Systems_with_AIML-Driven_Cybersecurity_in_ERP_Cloud_and_Snowflake_DB_Ecosystems/links/6798f322207c0c20fa62bf2c/Securing-Business-Intelligence-Systems-with-AI-ML-Driven-Cybersecurity-in-ERP-Cloud-and-Snowflake-DB-Ecosystems.pdf

20. Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. SN Computer Science, 2(5), 377.

    https://link.springer.com/article/10.1007/s42979-021-00765-8

21. Ahmad, A., Maynard, S. B., Desouza, K. C., Kotsias, J., Whitty, M. T., & Baskerville, R. L. (2021). How can organizations develop situation awareness for incident response: A case study of management practice. Computers & Security, 101, 102122.

    https://www.sciencedirect.com/science/article/abs/pii/S0167404820303953

22. Zohuri, B., Bowen, P. E., Kumar, A. A. D., & Moghaddam, M. (2022). Energy Driven by Internet of Things Analytics and Artificial Intelligence. J. Energy Power Eng., 16, 24-31.

    https://www.researchgate.net/profile/Akansha-Agarwal-6/publication/360073158_Energy_Driven_by_Internet_of_Things_Analytics_and_Artificial_Intelligence/links/6260416b8cb84a40ac7c7cb7/Energy-Driven-by-Internet-of-Things-Analytics-and-Artificial-

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

Intelligence.pdf

23. Shiva, R. (2022). The Role of AI in Securing Critical Infrastructure: A Data-Driven Approach to Cyber Defense.

    https://www.researchgate.net/profile/Ronaldo-Shiva/publication/388525378_The_Role_of_AI_in_Securing_Critical_Infrastructure_A_Data-Driven_Approach_to_Cyber_Defense/links/679bc49c4c479b26c9c2df46/The-Role-of-AI-in-Securing-Critical-Infrastructure-A-Data-Driven-Approach-to-Cyber-Defense.pdf

24. Vellani, K. H. (2019). Data-Driven Security. In Strategic Security Management (pp. 1-10). CRC Press.

    https://www.taylorfrancis.com/chapters/edit/10.4324/9780429506611-1/data-driven-security-karim-vellani

25. Zhu, Y., Zhang, Y., Wang, J., Song, W., Chu, C. C., & Liu, G. (2019, July). From data-driven to intelligent-driven: technology evolution of network security in big data era. In 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC) (Vol. 2, pp. 103-109). IEEE.

    https://ieeexplore.ieee.org/abstract/document/8754176

26. Okamoto, H. (2021). The Role of Information Security Event Management (SIEM) in Enhancing Intrusion Detection and Cybersecurity Through Machine Learning Technology.

    https://www.researchgate.net/profile/Hakim-Okamoto/publication/385085163_The_Role_of_Information_Security_Event_Management_SIEM_in_Enhancing_Intrusion_Detection_and_Cybersecurity_Through_Machine_Learning_Technology/links/6714ec99d796f96b8ec383de/The-Role-of-Information-Security-Event-Management-SIEM-in-Enhancing-Intrusion-Detection-and-Cybersecurity-Through-Machine-Learning-Technology.pdf

27. Girn, S. (2022, July). A data driven approach to board cybersecurity governance. In Pacific Asia Conference on Information Systems 2022. AIS.

    https://opus.lib.uts.edu.au/handle/10453/159073

28. Ivanova, S. (2022). Cybersecurity Challenges and Risk Management Strategies in Digital Sports Project Platforms. International Journal of Emerging Trends in Computer Science and Information Technology, 3(2), 23-31.

    https://www.ijetcsit.org/index.php/ijetcsit/article/view/63

29. Bechtsis, D., Tsolakis, N., Iakovou, E., & Vlachos, D. (2022). Data-driven secure, resilient and sustainable supply chains: gaps, opportunities, and a new generalised data sharing and data monetisation framework. International Journal of Production Research, 60(14), 4397-4417.

    https://www.tandfonline.com/doi/abs/10.1080/00207543.2021.1957506

30. Yin, J., Tang, M., Cao, J., You, M., & Wang, H. (2022). Cybersecurity applications in software: data-driven software vulnerability assessment and management. In Emerging trends in cybersecurity applications (pp. 371-389). Cham: Springer International Publishing.

    https://link.springer.com/chapter/10.1007/978-3-031-09640-2_17

31. Newell, A. (2021). Optimizing Cloud Infrastructure: AI/ML Solutions for Snowflake Databases and Business Intelligence.

    https://www.researchgate.net/profile/Allen-Newell/publication/387996057_Optimizing_Cloud_Infrastructure_AIML_Solutions_for_Snowflake_D atabases_and_Business_Intelligence/links/6786360e55274940f1263018/Optimizing-Cloud-Infrastructure-AI-ML-Solutions-for-Snowflake-Databases-and-Business-Intelligence.pdf

32. Reddy, P. S., & Pelletier, J. M. (2022, May). The pentest method for business intelligence. In 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO) (pp. 1117-1125). IEEE.

    https://ieeexplore.ieee.org/abstract/document/9803788

33. Torres, M. A. E., Guerrero, F. T., & Budgud, A. T. (2022, December). Data-Driven Cyber Threat Intelligence: A Survey of Mexican Territory. In 2nd EAI International Conference on Smart Technology (pp. 89-110). Cham: Springer International Publishing.

    https://link.springer.com/chapter/10.1007/978-3-031-07670-1_7

34. Andreassen, J., Eileraas, M., Herrera, L. C., & Noori, N. S. (2022, October). Increase: A dynamic framework towards enhancing situational awareness in cyber incident response. In International Conference on Information Technology in Disaster Risk Reduction (pp. 230-243). Cham: Springer Nature Switzerland.

    https://link.springer.com/chapter/10.1007/978-3-031-34207-3_15

35. AlSadhan, T., & Park, J. S. (2021, December). Leveraging information security continuous monitoring to enhance cybersecurity. In 2021 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 753-759). IEEE.

    https://ieeexplore.ieee.org/abstract/document/9799002'

36. Tseng, M. L., Bui, T. D., Lim, M. K., Fujii, M., & Mishra, U. (2022). Assessing data-driven sustainable supply chain management indicators for the textile industry under industrial disruption and ambidexterity. International Journal of Production Economics, 245, 108401.

    https://www.sciencedirect.com/science/article/abs/pii/S0925527321003777

37. Lee, C. S., Cheang, P. Y. S., & Moslehpour, M. (2022). Predictive analytics in business analytics: decision tree. Advances in Decision Sciences, 26(1), 1-29.

    https://www.proquest.com/openview/3453584715adbe9094f8bd061f67f64d/1?pq-origsite=gscholar&cbl=25336

38. Bachmann, N., Tripathi, S., Brunner, M., & Jodlbauer, H. (2022). The contribution of data-driven technologies in achieving the sustainable development goals. Sustainability, 14(5), 2497.

    https://www.mdpi.com/2071-1050/14/5/2497

39. Crotty, J., & Daniel, E. (2022). Cyber threat: its origins and consequence and the use of qualitative and quantitative methods in cyber risk assessment. Applied Computing and Informatics, (ahead-of-print).

    https://www.emerald.com/insight/content/doi/10.1108/aci-07-2022-0178/full/html

40. Alonso, G. (2022). Leveraging Snowflake DB for AI/ML-Driven Business Intelligence and Cybersecurity in ERP Cloud Systems.

    https://www.researchgate.net/profile/Gustavo-Alonso-6/publication/388452750_Leveraging_Snowflake_DB_for_AIML-Driven_Business_Intelligence_and_Cybersecurity_in_ERP_Cloud_Systems/links/6798fc868311ce680c3f1421/Leveraging-Snowflake-DB-for-AI-ML-Driven-Business-Intelligence-and-Cybersecurity-in-ERP-Cloud-Systems.pdf

41. Ahsan, M., Nygard, K. E., Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022). Cybersecurity threats and their mitigation approaches using Machine Learning—A Review. Journal of Cybersecurity and Privacy, 2(3), 527-555.

    https://www.mdpi.com/2624-800X/2/3/27

42. Andrade, R. O., & Yoo, S. G. (2019). Cognitive security: A comprehensive study of cognitive science in cybersecurity. Journal of Information Security and Applications, 48, 102352.

    https://www.sciencedirect.com/science/article/abs/pii/S2214212618307804

43. Tambare, P., Meshram, C., Lee, C. C., Ramteke, R. J., & Imoize, A. L. (2021). Performance measurement system and quality management in data-driven Industry 4.0: A review. Sensors, 22(1), 224.

    httpswww.mdpi.com/1424-8220/22/1/224

44. Kenzie, F. (2021). Integrating Artificial Intelligence with Database Technologies: A New Frontier in Cybersecurity.

    https://www.researchgate.net/profile/Florence-Kenzie/publication/385410698_IntegrKara, M. E., Fırat, S. Ü. O., & Ghadge, A. (2020). A data mining-based framework for supply chain risk management. Computers & Industrial Engineering, 139, 105570.ating_Artificial_Intelligence_with_Database_Technologies_A_New_Frontier_in_Cybersecurity/links/67235415db208342dee09ba5/Integrating-Artificial-Intelligence-with-Database-Technologies-A-New-Frontier-in-Cybersecurity.pdf

45. Wang, K., Guo, X., & Yang, D. (2022). Research on the effectiveness of cyber security awareness in ICS risk assessment frameworks. Electronics, 11(10), 1659.

    https://www.mdpi.com/2079-9292/11/10/1659

46. Kara, M. E., Fırat, S. Ü. O., & Ghadge, A. (2020). A data mining-based framework for supply chain risk management. Computers & Industrial Engineering, 139, 105570.

    https://www.sciencedirect.com/science/article/abs/pii/S0360835218306156

47. Olayinka, O. H. (2022). Ethical implications and governance of AI models in business analytics and data science applications. International Journal of Engineering Technology Research & Management.

    https://www.researchgate.net/profile/Olayinka-Olalekan/publication/390348655_ETHICAL_IMPLICATIONS_AND_GOVERNANCE_OF_AI_MODELS_IN_BUSINESS_ANALYTICS_AND_DATA_SCIENCE_APPLICATIONS/links/67edc90a03b8d7280e1e20dd/ETHICAL-IMPLICATIONS-AND-GOVERNANCE-OF-AI-MODELS-IN-BUSINESS-ANALYTICS-AND-DATA-SCIENCE-APPLICATIONS.pdf

48. Mihailescu, M. I., & Nita, S. L. (2022, September). Towards Data Science for Cybersecurity: Machine Learning Advances as Glowing Perspective. In Proceedings of SAI Intelligent Systems Conference (pp. 26-48). Cham: Springer International Publishing.

    https://link.springer.com/chapter/10.1007/978-3-031-16078-3_2

49. Owen, A., & Ajeigbe, K. (2021). Addressing Cybersecurity in AI-Enhanced Manufacturing Systems.

    https://www.researchgate.net/profile/Kolade-Ajeigbe-2/publication/390366806_Addressing_Cybersecurity_in_AI-Enhanced_Manufacturing_Systems/links/67ebd0e476d4923a1aeb7011/Addressing-Cybersecurity-in-

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

AI-Enhanced-Manufacturing-Systems.pdf

50. Ali, F. (2021). Revolutionizing Cloud Computing with AI/ML for Business Intelligence, ERP Cloud, and Snowflake DB Security Enhancements.

    https://www.researchgate.net/profile/Farman-Ali-41/publication/388448709_Revolutionizing_Cloud_Computing_with_AIML_for_Business_Intelligence_ERP_Cloud_and_Snowflake_DB_Security_Enhancements/links/6798bf6b8311ce680c3ecb1a/Revolutionizing-Cloud-Computing-with-AI-ML-for-Business-Intelligence-ERP-Cloud-and-Snowflake-DB-Security-Enhancements.pdf

51. Bitomsky, L., Bürger, O., Häckel, B., & Töppel, J. (2020). Value of data meets IT security–assessing IT security risks in data-driven value chains. Electronic Markets, 30, 589-605.

    https://link.springer.com/article/10.1007/s12525-019-00383-6

52. Treacy, S. (2022). Ensuring compliance in the digital era: A knowledge-based dynamic capabilities framework wheel for data-driven organisations. International Journal of Business Analytics and Intelligence, 10(2), 25.

    https://www.proquest.com/openview/407d7e39071535c9ec9fbf381bc5adba/1?pq-origsite=gscholar&cbl=2043514

53. Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2021). A review of data-driven decision-making methods for industry 4.0 maintenance applications. Electronics, 10(7), 828.

    https://www.mdpi.com/2079-9292/10/7/828

54. Neshenko, N. (2021). Illuminating Cyber Threats for Smart Cities: A Data-Driven Approach for Cyber Attack Detection with Visual Capabilities (Doctoral dissertation, Florida Atlantic University).

    https://www.proquest.com/openview/344d68ce2e79f1b35824f71a1e688158/1?pq-origsite=gscholar&cbl=18750&diss=y

55. Debar, H. Security Operations & Incident Management Knowledge Area Issue.

    https://smiitcyberai.com/resources/docs/Security_Operations__Incident_Management_issue_1.0.pdf

56. Strohmeier, M., Pavur, J., Martinovic, I., & Lenders, V. (2021). Studying neutrality in cyber-space: a comparative geographical analysis of honeypot responses. In Critical Information Infrastructures Security: 16th International Conference, CRITIS 2021, Lausanne, Switzerland, September 27–29, 2021, Revised Selected Papers 16 (pp. 186-203). Springer International Publishing.

    https://link.springer.com/chapter/10.1007/978-3-030-93200-8_11

57. Lenders, V. (2022). Studying Neutrality in Cyber-Space: a Comparative Geographical Analysis of Honeypot Responses. In Critical Information Infrastructures Security: 16th International Conference, CRITIS 2021, Lausanne, Switzerland, September 27–29, 2021, Revised Selected Papers (Vol. 13139, p. 186). Springer Nature.

58. Samtani, S., Abate, M., Benjamin, V., & Li, W. (2019). Cybersecurity as an industry: A cyber threat intelligence perspective. In The Palgrave Handbook of International Cybercrime and Cyberdeviance (pp. 1-20). Palgrave Macmillan, Cham.

    https://link.springer.com/rwe/10.1007/978-3-319-90307-1_8-1

59. Koroniotis, N., Moustafa, N., Schiliro, F., Gauravaram, P., & Janicke, H. (2020). A holistic review of cybersecurity and reliability perspectives in smart airports. IEEE Access, 8, 209802-209834.

ISSN 2576-5973 (online), Published by "Global Research Network LLC"
under Volume: 6 Issue: 8 in Aug-2023 https://globalresearchnetwork.us/index.php/ajebm

https://ieeexplore.ieee.org/abstract/document/9252856

60. Bhardwaj, A. (2021). Cybersecurity incident response against advanced persistent threats (APTs). Security Incidents & Response Against Cyber Attacks, 177-197.

    https://link.springer.com/chapter/10.1007/978-3-030-69174-5_9

61. Boeding, M., Boswell, K., Hempel, M., Sharif, H., Lopez Jr, J., & Perumalla, K. (2022). Survey of cybersecurity governance, threats, and countermeasures for the power grid. Energies, 15(22), 8692.

    https://www.mdpi.com/1996-1073/15/22/8692

62. Wang, C., & Zhu, H. (2022). Wrongdoing monitor: A graph-based behavioral anomaly detection in cyber security. IEEE Transactions on Information Forensics and Security, 17, 2703-2718.

    https://ieeexplore.ieee.org/abstract/document/9830760

63. Solomon, A., Michaelshvili, M., Bitton, R., Shapira, B., Rokach, L., Puzis, R., & Shabtai, A. (2022). Contextual security awareness: A context-based approach for assessing the security awareness of users. Knowledge-Based Systems, 246, 108709.

    https://www.sciencedirect.com/science/article/abs/pii/S0950705122003276

64. Dataset Link: https://www.kaggle.com/datasets/saurabhshahane/honeypot-15