Multiple sources of big data are used to create a method for protecting computer networks

Mohammad Eid Alzahrani

Department of Computer Science, Faculty of Computing & Information
Al-Baha University, Al-Baha, Saudi Arabia
meid@bu.edu.sa

Abstract: The purpose of this article is to present a model for the cybersecurity defense of computer networks that makes use of big data from multiple sources. The purpose of this endeavor is to improve the overall security of computer networks by addressing the limitations of the defense systems that are currently in place. A comprehensive analysis of the current state of network security is carried out, with a particular emphasis placed on the difficulties that are encountered in this field. After that, the concept of big data that comes from multiple sources is presented as a potential solution. A definition of big data and an analysis of the multisource big data model are presented in this article. An information system network security framework is presented that can be found in this article. The model illustrates the connection between network operations, potential security risks, attacks on networks, and the defense provided by security devices. For the purpose of developing a defense system measurement and optimization system, the network security system measurement and optimization scheme is utilized. Real-world scenarios are skillfully incorporated into the application analysis that is being conducted for the project. The purpose of this article is to demonstrate the usefulness and efficiency of the proposed network security defense system evaluation and optimization scheme. This is accomplished by evaluating and enhancing the security defense system through the utilization of conventional methods.

Keywords: *Big data, Multiple Sources, Computer Network Security, Cyberattacks, Network Protection.*

استخدام مصادر متعددة للبيانات الضخمة لإنشاء طريقة لحماية شبكات الكمبيوتر

الملخص: الهدف من هذه الورقة العلمية هو تقديم نموذج لتعزيز الامن السيبراني لشبكات الحاسب وذلك بالاستفادة واستخدام البيانات الضخمة من مصادر متعددة. والغرض من هذا المسعى هو تحسين الأمان العام لشبكات الكمبيوتر من خلال معالجة القيود الموجودة في أنظمة التعزيز الامني الحالية. يتم إجراء تحليل شامل للحالة الحالية لأمن الشبكات، مع التركيز بشكل خاص على الصعوبات التي يتم مواجهتها في هذا المجال. بعد ذلك، يتم طرح مفهوم البيانات الضخمة القادمة من مصادر متعددة كحل محتمل. تستعرض الورقة تعريف للبيانات الضخمة وتحليل لنموذج البيانات الضخمة متعدد المصادر. يتم عرض إطار عمل لأمن شبكة نظم المعلومات يمكن العثور عليه من خلال هذا البحث. النموذج يوضح العلاقة بين عمليات الشبكة، المخاطر الأمنية المحتملة، الهجمات على الشبكات، والتعزيز الأمني التي توفرها أجهزة الخاصة بالأمان. ومن أجل تطوير نظام قياس وتحسين التعزيز الأمني، يتم استخدام مخطط قياس وتحسين نظام الأمان للشبكات الحاسوبية وذلك من خلال دمج السيناريوهات الواقعية بشكل فعال في تحليل التطبيق الذي يتم إجراؤه للمشروع. الغرض من هذا المقال هو إظهار فائدة وكفاءة خطة تقييم وتحسين نظام التعزيز الأمني المقترحة. يتم تحقيق ذلك من خلال تقييم وتعزيز نظام الحماية الأمني باستخدام الطرق التقليدية.

1. Introduction

Recently, there has been a significant increase in the progress of emerging technologies such as blockchain (Zhou, Z. 2022), the Internet of Things (IoT), cloud computing resources (Zarei S.M., 2021), and big data. The integration of computer networks as the underlying framework for information construction has had a significant influence on both economic progress and human ways of life (Prvan, M, 2020).

The number of connected devices and generated content are growing rapidly on the internet. While networks offer various conveniences, the possibility of attacks can give rise to security apprehensions (Nour, B, 2021). The complete utilization of multi-source and large data resources can be achieved through the mining of explorer travel rules and the acquisition of trip information through big data advancement (A. Ju, 2020). Additionally, monstrous learning benefits from the consistent absorption of vast quantities of test data and the separation of sporadic, unsteady, and exceptionally dubious data ascribes. Additionally, it can provide state-of-the-art development and advancement to address the accuracy of explorer stream assumption in various environments in metropolitan rail travel voyager stream assumption data support. While considering the impact of social and monetary components, temporary and spatial factors, and various emergencies on explorer stream changes, these variables can be considered continuously and exhaustively (Bhat, 2021).

In the contemporary mechanical and associated world, it is imperative to protect computer networks from a variety of cyber threats. A significant instrument for reinforcing these networks has emerged as a result of the remarkable development of data generated from a variety of sources: big data. The utilization of various big data sources, which are areas of strength, can be employed to establish a comprehensive strategy for safeguarding computer networks from emerging cyber threats (C. Zhou, 2021). Network traffic is a significant source of big data. The identification of unusual behaviors that are indicative of potential hazards can be facilitated by the examination of significant surges in network sections, which can be used to establish standard examples of conduct.

Additionally, network device, server, and application logs provide valuable insights into organizational practices by facilitating the identification of unauthorized access attempts and questionable behavior (D. Wang, 2020). In essence, organizations have the ability to implement a variety of strategies to protect computer networks by leveraging a limited number of large data sources. By integrating network traffic analysis, structure logs, threat knowledge, and IoT data, this approach establishes a comprehensive security system.

This concept aids organizations in maintaining a competitive edge over cyber adversaries and safeguarding their critical high-level assets by means of continuous monitoring, analysis, and proactive risk mitigation (Fadhil, 2021).

2. Literature review

The network produces a daily volume of 2.5 exabytes of data, which is substantial, diverse, and generated quickly (R. Vinayakumar, 2017). As big data is collected and analyzed more efficiently, the significant value that is concealed within the data is gradually being uncovered. Network operators can optimize network performance and enhance network revenue by leveraging big data (H. Sun, 2021). The article by Gupta (2020) provides a meticulous logical categorization and risk assessment for artificial intelligence models used in secure data analysis. The paper resolves the central inquiry of guaranteeing data security and protection in the period of artificial intelligence driven examination. The creators order the different man-made intelligence models utilized in data examination through a careful survey of past writing, and afterward recommend a risk model to recognize possible weaknesses. The review gives significant experiences to specialists and experts to foster solid security apparatuses for data examination structures via cautiously arranging simulated intelligence models and related gambles.

This paper conducts a deliberate examination that focuses on the intersection of IoT advancements and big data in clever settings (Hajjaji, 2021). In particular, the review examines the intersection of IoT applications and big data analysis, particularly in relation to smart metropolitan regions, medical care, transportation, and energy for Presidents. The creators provide pieces of information regarding the current state of craftsmanship, challenges, and future prospects in this thriving region by combining findings from a variety of investigations. The deliberate review showcases the diverse applications of big data and IoT advancements, ranging from continuous observation and vision investigation to custom-made services and resource enhancement. In addition, the review emphasizes the necessity of addressing critical issues such as data security, adaptability, and interoperability in order to gain a comprehensive understanding of the potential impact of big data and IoT on the development of sharp ecosystems.

Data fusion brought a fresh approach for the identification of heterogeneous intrusions (Jeyepalan and Kirubakaran, 2019). (Essid and Jemili, 2016) proposed using Hadoop and MapReduce the integration of two heterogeneous data sources. Combining intrusion detection datasets—including the NSL-KDD, Mawilab, and DARPA'99 datasets—Ben Fekih and Jemili, 2018 put forth a method They built and assessed the detection model using the Naïve Bayes algorithm. In order to suggest an approach for the discovery of temporal patterns, (Radhakrishna V et al., 2019) presented the idea of data fusion in respect to the temporal pattern tree. Every timeslot generates a tree; the trees acquired for one timeslot are combined or fused to produce the total tree for the whole dataset. Effective and proactive pruning of elements during the pattern mining process depends much on the idea of tree-based data fusion. By means of principal component analysis (PCA), (Om Prakash Singh et al., 2022) sought the suitable coefficients for data fusion. Look at how computational information approaches are utilized in numerous spaces for big data examination (Iqbal, 2020).

The review gives a complete outline of a few computational experiences, like brain networks, hereditary calculations, fluffy rationale, and multiverse information, among others. The creators outline the adequacy of these strategies in tending to perplexing difficulties related with large data examination, for example, data pre-taking care of, feature determination, and model affirmation, through relevant investigations and true applications.

The study by J. Hu (2021) provides a methodology for predicting problems associated with vehicle travel by utilizing data from various sources. The survey focuses on accurately predicting protests by utilizing various data inputs, such as GPS data, traffic patterns, and real-time travel information. The recommended approach aims to enhance the accuracy of genuine assumption by integrating diverse data sources and implementing artificial intelligence techniques. This has significant implications for efficient urban driving, navigation systems, and transportation planning. This examination aims to enhance intelligent transportation systems by limiting the capacity of multi-source data analysis for advanced research.

The study conducted by Khang and colleagues in 2024 Examine the potential applications of big data in resolving challenges within the pharmaceutical industry (Khang, 2024). Utilizing a vast amount of patient data, clinical records, and genetic information, big data research enables medical professionals to identify significant new discoveries, customize treatments, and focus on long-term outcomes. This segment showcases the profound impact of big data on shaping the future of medicine and introduces several applications, such as drug discovery, precision medicine, and predictive market analysis. This examination contributes to the growing body of literature on the application of data-driven approaches to enhance clinical development and healthcare practices.

Wang et al. (2020) propose a pre-impact fall area framework for a CNN enterprise that incorporates multiple sources (L. Wang, 2020). The survey addresses the primary requirement for accurate fall detection, particularly in vulnerable populations such as the elderly. The proposed framework achieves precise and robust fall detection prior to impact by integrating data from multiple sensors, such as depth cameras, gyroscopes, and accelerometers.

The dress showcase strategy by Tehrany et al. (2019) predicts tropical woodland fire weakness spatially. Forest fires pose significant financial and environmental risks in tropical regions, so the audit addresses the growing need for precise assessment. With the Logit Lift computer-based intelligence classifier and many geospatial data sources, the designers created an impressive prescient model that can identify forest fire-prone areas. The study improves proactive fire fighting and early warning systems, reducing the harm caused by wildfires. Sanden and Neideck (2021) examine how connected data resources can improve multi-source public region tirelessness. The survey uses public data sources to create complete, connected data resources for evidence-based policymaking, program assessment, and organization delivery. The examination promotes data executive techniques and government organization cooperation to make data an incentive for social benefit.

The cybersecurity improvements at specific southwest Nigerian schools are examined by Oluwafunmilayo (2019). The review evaluates academic institutions' cybersecurity measures to protect sensitive data and cyber threats. The findings highlight the importance of cybersecurity availability and protecting stakes in strong security measures to reduce cyber risks. Analyse large data using data science and computerized reasoning (Pramanik, 2023). This section describes big data research's challenges and opportunities. The authors explain big data analysis's origins and potential for independent direction, progress, and social change by examining various keen strategies, instruments, and application spaces. The examination improves big data strategy understanding and multidisciplinary collaborative efforts to maximize big data's potential in many areas.

An proposed method for intrusion detection focuses on combining data from different sources, like user behavior, system logs, and network traffic (Anjum, N. et al., 2021). This method is specifically made to make intrusion detection more accurate. Aleroud and Karabatis (2017) suggested a context-aware data fusion method that makes intrusion detection more accurate by taking into account things like the time and location of the intrusion.

3. Design of a multisource big data based computer network security defence system

3.1.Building Multisource Big Data Models

Providing clients with information regarding data plan and various perspectives through a unified perspective is the primary objective of the multisource big data organizing stage. Consequently, the data availability process is further developed by enhancing the data organizing cycle's instinct.

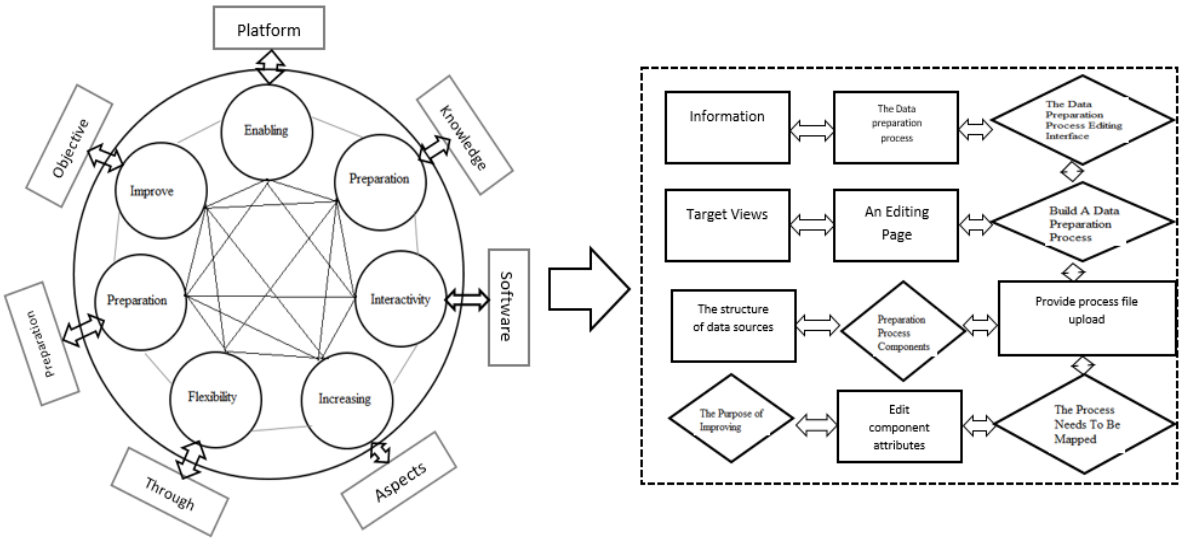


Figure 1: Process of the platform functional analysis system.

Information regarding the development of data sources and target views must be accumulated in order to achieve the objective of disseminating a unified perspective. It is also necessary to provide a modifying page that allows clients to select specific sections of the data availability cycle and modify the credit allocations in the data organizing process by adjusting the association point. Finally, to satisfy the justification for managing the versatility of the data organization process, a data organization process for multisource big data must be established, as well as process record move and age capabilities provided. The cycle should be developed and demonstrated after the framework is established, and it is typically employed to finalize the client-depicted strategy. Each step toward the client-represented data game plan process should be executed with adaptability and efficiency (Q. Guo, 2020). It is imperative to enhance the estimation and strategy for addressing missing data. The bit by bit advantageous assessment system approach is illustrated in Figure 1.

Figure 2 shows the multisource big data flowchart. Extra-assembled analysis of multisource security events generates anomaly alerts using security semantics. Though assault examinations can be fooled, anomaly alerts provide only low-level security. Relationship analysis will link these spots to a more serious attack scenario using semantic security data from firewalls, antivirus software, and intrusion detection systems (Ragazou, 2023).

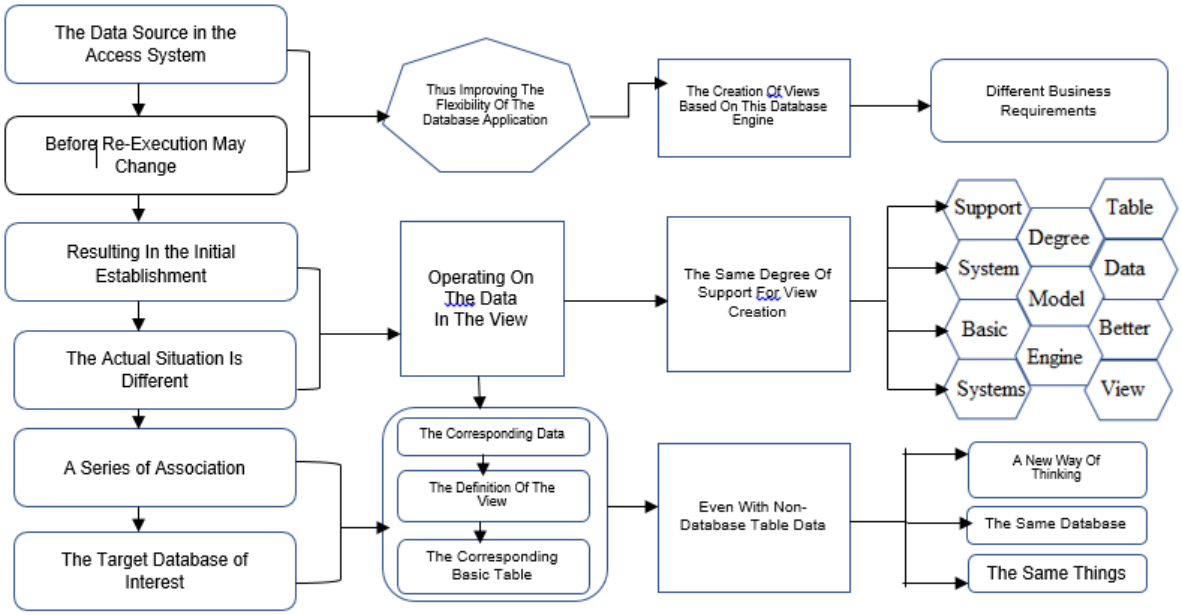


Figure 2: Big data flowchart with multiple sources.

3.2.System Model Design is Defined by Computer Network Security

The goal of the quantitative evaluation of defense systems during the arrangement stage is to help security staff figure out if security threats are built into the system and, if so, which threats the system is meant to protect against. It also wants to find any holes in the defense system or threats that the system might not be able to stop well enough (S. S. Harsha, 2019).

(1) Figure out the possible security risks and set up a framework for the evaluation subject's network security. There is a suggested network security model for the information system that shows the network topology of the information infrastructure.

Not all of the device hubs that the information structure goes through are completely safe, which makes it less safe (Tao, 2020).

(2) The degree of risk associated with security threats is used to assess the system's evolutionary stage. Various research procedures are employed to assess the severity of a security risk. Information systems may be vulnerable to upcoming challenges, such as security risks. When information systems are threatened by attacks, security personnel counteract these threats by deploying security devices to neutralize or eliminate them (Ullah, 2023). This suggests that the degree of probability that the information system assesses can be employed as a quantitative indicator to measure the protective capability of the shield structure. This article examines the dynamic relationship between the type of danger and the actual level of risk. It utilizes various reviewed studies to assess the level of risk associated with a specific situation. Aggressors interpret dangers as clear attacks and achieve their goals by employing hostile behaviors to deliver threats to their intended targets. Risks associated with security encompass data leakage, data aggregation, data manipulation, and lack of managerial accessibility. Specific hazards present distinct security opportunities and vary in the level of risk they pose to a security threat T at a device hub within a data infrastructure (Vasa, 2023). Using SDNA design, a hypervisor is put between each network hub. This gives each hub a unique look while still being essential for the computer's OS, different applications, and end consumers.

$$\sum_{i=1} \boldsymbol{tw}(\boldsymbol{n}-\mathbf{1}) = \frac{\sqrt{\boldsymbol{n}-\mathbf{1}}}{\boldsymbol{n}+\mathbf{1}} \qquad (\mathbf{1})$$

(3) Utilize the network security model to decide how to safeguard the information structure's security gadgets against dangers to the gadget centers. There is a security risk since it is normal that the contraption center point is gadget 1. T, device1's network danger is addr1, the assailant's network address is addr2, and the security gadget's defensive impact is contraption against the danger in the shield system. DTj (dj1, dj2,... djn), dji with ti adjusted correspondence, addresses T. Clients will work with computer information systems for a lot of time consistently (Wu, 2020). Bosses frequently start with a model setup of the security of the computer system, which safeguards the computer information structure:

$$\boldsymbol{S} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} (\boldsymbol{r_i} + \boldsymbol{tw_{i-1}}) - \boldsymbol{TW^T}. \qquad (\mathbf{2})$$

(4) Decide if the watchman design is suitable for general protection. Considering that the information structure has m resources, the resources' genuine monetary worth is utilized to decide their significance (X. Chen, 2019).

The asset importance weight is communicated as AW (aw1,... aw2), where aw1: aw2:...: awm v1: v2:...: vm, where v addresses the genuine money related worth of each buy, and j1awj, the equation for deciding the protection effect of the all-out information structure is acquired:

$$S_1 = \sum_{j=1}^{m} aw_j + \sum_{i=1} tw_i. \qquad (3)$$

The data sources can be recognized, recognized, and dealt with fittingly by the security protect engineering in view of the aftereffects of data coordinating.

Figure 3 shows the security defend system's design. The planning module, the disclosure module, the actually take a look at module, and the database storing system make up the shield part. The structure will initially gather a large measure of data (T1, T2, TN, and so forth) in the planning module. It will then pre-process the data, store the pre-dealt with data in the readiness set database, set up the pre-taken care of planning set utilizing the mind net, and store the got features the component library (Zhang, 2021).

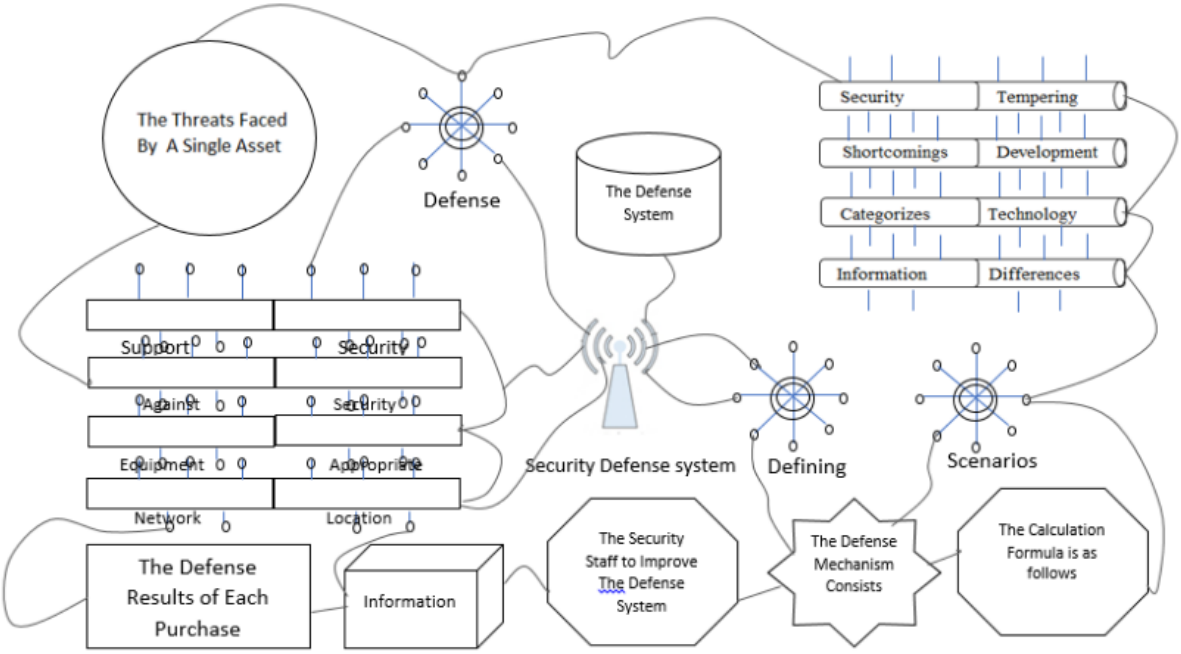


Figure 3: The security defense system's design.

At last, two situational points of view are analyzed in line with their reasonable relevance. The first is RE, the general reaction impact of the protection system against the hazards that gadget 1 poses in the resources; re1 0 shows that gadget 1 is covered by the defend range of the protection structure. All security devices cannot differentiate the danger presented by attacks; hence, none of the security devices covered by the defense system (gadget 1) can prevent attacks linked with hazards; moreover, none of the security gadgets covered by the protect structure (gadget 3) can prevent attacks linked with hazards. Therefore, assuming each of the four credits of gadget 1 are available, no security element can protect against the risk presented by re4 0.

The next one is Bij, Bij, the reaction system of the security devices to protect the resources against the risk gadget 1. Though every value device 1 has a similar significance as the others, Bij, or the actual response to hazards that every remarkable security gadget generates, really counts (Z. Xiong, 2021).

3.3 Potential Situations anomaly notifications using security semantics

Anomaly correction for data quality is essential for improving datasets' accuracy, reliability, and usability across all domains and use cases. We will highlight the most important areas where our framework can improve data quality in the sections that follow.

Personal details: The accuracy of sensitive data can be compromised due to human error, incomplete or missing data, inconsistent data formats or types, misspellings, and other human error-related issues. There is hope that the proposed framework can improve data quality. Customer profiling, targeted marketing, and personalized services are all at risk when data is inaccurate or unreliable due to quality anomalies. Some methods for dealing with these problems include finding correlated features, choosing comparable records, and filling in missing values, fixing inconsistencies, and fixing data entry mistakes with machine learning models like XGBoost. Think about a scenario where the date of birth field has some blanks. To improve the data quality and make accurate predictions of date of birth values, the framework can use correlated features like name, address, and age. (Elouataoui Widad, 2023).

Cybersecurity : When it comes to cybersecurity, data quality correction is absolutely crucial. In order to effectively detect threats, respond to incidents, and manage risks, trustworthy cybersecurity data is essential. Our dependence on digital systems and the ever-changing threat landscape make this a top priority (Gahi, Y., 2019). When it comes to cybersecurity, poor data quality can be caused by things like missing or incomplete log entries, incorrect timestamps, inconsistent data formats, and intrusion detection systems that produce false positives or false negatives. The cybersecurity dataset data quality could be improved by the framework.

Healthcare: Inaccurate patient records, inconsistent or missing diagnoses, and treatment discrepancies are some causes of poor data quality in healthcare information systems. The reliability of medical research, decisions, and patient care can all be severely compromised by these anomalies. These problems can be handled by the framework by selecting related features like patient IDs, medical diagnoses, and treatment histories.

Trasportation: Anomalies in data quality can occur in transportation networks due to things like incorrect vehicle identification numbers, missing route information, inconsistent timestamps, and inaccurate location data. Errors can make it harder to find the best route, predict delays, and control traffic.

These problems can be handled by the framework by picking out related features like location coordinates, timestamps, and vehicle attributes.
Banking : In the banking industry, inconsistent transaction records, erroneous customer information, missing account details, and inconsistencies in financial statements can all result from poor data quality. Problems with noncompliance, ill-informed decisions, and financial losses can all stem from aberrations. To solve these problems, the framework can identify and examine interrelated client identifiers, account details, and transaction types. By using anomaly detection and filling in missing or inconsistent data, the framework can validate financial datasets, fix transaction errors, and reconcile account information. Financial institutions, transportation companies, healthcare providers, online retailers, and personal information managers can all benefit from the data quality anomaly correction framework.

4. Results analysis

4.1 Examination Analysis of a comprehensive data model using multiple sources

The system optimizes and expands standard processes to address issues in multisource significant data unification, taking into account the data preparation requirements of multisource analysis and centralization management. The solution pertains to the establishment of the view, the definition of data preparation process components, and their organization. Computer networks provide convenience and enhanced office efficiency; however, they also present security risks, including software vulnerabilities, hacker attacks, malicious codes, and protocol vulnerabilities. These threats have the potential to inflict varying degrees of damage on computers.

The multisource human data platform's data preparation results are saved and supported for analysis and centralized management tasks. Enabling unified big data management from access to output and displays, the e-platform integrates human data access, editing process management, and execution. Data is saved as part of the data preparation process to facilitate subsequent tasks and increase flexibility in multisource big data. The attacker is presented with a significant obstacle by the weakest attack link, as they are required to generate events at various stages of the attack chain in order to accomplish their goal.

However, attack independence implies that there is little chance of attack chain related events happening in the absence of an attack. Thus, it follows that the coincidence of multiple events in an attack chain is implied. On the other hand, analysts can reconstruct the attack scenario by using the association between these events. This article also focuses on using different source events for association analysis. The analysis presented suggests that studies concentrating on network attack chains have a clear advantage when it comes to characterizing attack scenarios and improving comprehension. Furthermore, it is possible to recognize crucial phases in the attack implementation process and quickly intercept them during defense by using forwarding mapping and reverse reasoning of attack chains.

The analysis and detection of Advanced Persistent Threats (APTs) frequently lag behind the actual attack due to the slowness of defense mechanisms. The dynamic security model suggests that there may be a longer lag between the time an attack is discovered and successfully carried out. Et-Dt + Rt-Pt > 0. is the equation. Only when Pt is greater than Dt + Rt can the system be guaranteed to be secure.

The MCKC network attack chain model can be used to decrease the time it takes to detect an attack, thereby increasing system security. The relationship's description cannot exist without the keyword. The variable n(t, R) indicates how frequently it occurs in R. The following is the equation to calculate P(t|θees). To create a single, cohesive viewpoint, information about the target views' and data sources' structures must be gathered.

The flexibility of the data preparation process should be increased by including an editing page that is easy to use. On this page, users will be able to choose which components of the data preparation process to edit and change the properties of those components. In addition, functions for uploading and generating process files should be included in a data preparation process that can handle multiple big data sources.

The MCKC model improves on earlier techniques in most metrics. The lightweight model retains a recursive structure to faithfully represent the lateral movement within the internal network while condensing the attack process into five stages. Because the bidirectional analysis method supports both cyclic iterative analysis and metadata analysis, it is more akin to human analytical cognition.

One advantage of the proposed MCKC model is that it can help analysts who analyze and comprehend complex attack events in large enterprise networks by reducing their cognitive load. This is achieved by applying a reasoning process that is similar to that of a human analyst and merging data from multiple sources. In order to address the scalability issue during the analysis process, this approach also makes it easier to integrate additional attack processes and creative probes into traditional network environments (Zhou, Z, 2022). Nevertheless, a crucial aspect of APT attack research is quantitative analysis, which the MCKC model lacks. Attack operations need to use tactics that make their actions difficult for defenders to detect.

4.2 Solution Implementation is Defined by Computer Network Security

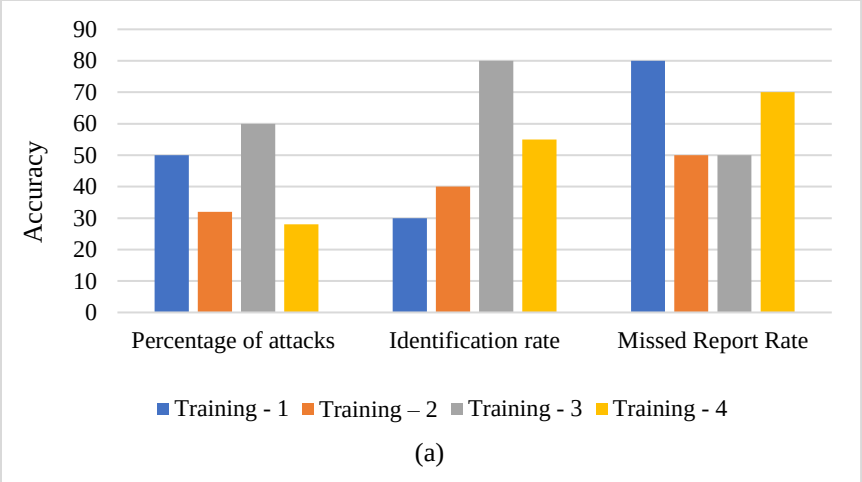
Many methods, including encryption, authentication, identification, and access control, are included in security defense systems. Management is the methodical application of particular security resources to achieve a specific goal.

A security defense system's management involves a variety of activities, including risk assessment, planning, the acquisition of systems and services, authentication, maintenance, and the establishment of policies, standards, and procedures. The information assurance program is significantly influenced by the contributions of individuals, and security defense systems encompass both personnel security and security personnel.

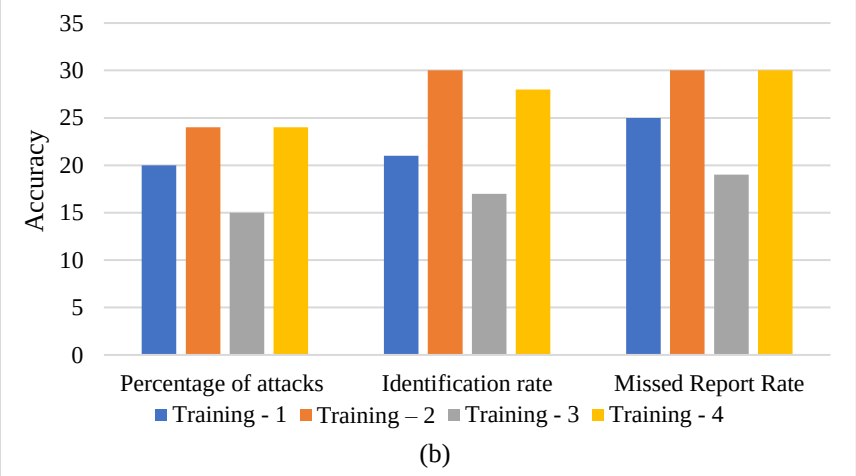
Individuals must possess a high level of security consciousness, knowledge, and proficiency in safety protocols in order to properly design, implement, and supervise security measures. Computer administrators frequently perpetrate specific attacks that originate within the organization. Consequently, it is imperative to possess a comprehensive comprehension of security defense systems. The computer information systems will be frequently used and the users will engage in prolonged operation. In computer systems, administrators typically commence with a model configuration of the system's security, which offers a specific level of protection for the computer information system. Nevertheless, the computer system is unable to accurately identify and prevent malicious attacks, spamming, SQL injection, and improper user operations.

The administration of the system is significantly impeded by these malevolent operations. Based on four critical factors: spamming, SQL injection, user behavior analysis, and access address, we offer a thorough evaluation and analysis. Figure 4 illustrates the degree of user accuracy variability.

Comparative tests were used to assess the impact of the number of iterations on the training set's accuracy, the validation set's accuracy difference, the training set's loss function value change, and the validation set's loss function value change. Additionally, each dataset's various classifications' accuracy, recall, and f1-score were assessed and examined. The experiments compare and assess the conventional approaches to classification: multilayer perceptrons, recurrent neural networks, and basic Bayesian techniques.


90
80
70
60
50
40
30
20
10
0
Accuracy
Percentage of attacks
Identification rate
Missed Report Rate
Training - 1
Training – 2
Training - 3
Training - 4

(a)


35
30
25
20
15
10
5
0
Accuracy
Percentage of attacks
Identification rate
Missed Report Rate
Training - 1
Training – 2
Training - 3
Training - 4

(b)

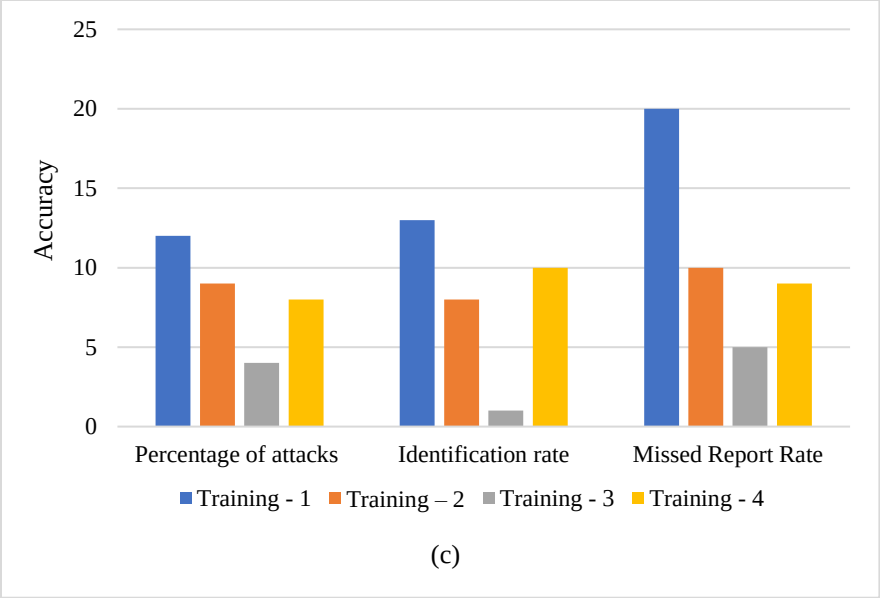


Figure 4: User behaviour accuracy.

Comparative tests were conducted on the exploratory data to adjust the precision of the preparation set in relation to the number of emphasis, the exactness of the approval set in relation to the number of cycles, the worth of misfortune capability of the preparation set in relation to the number of cycles, and the worth of misfortune capability of the approval set in relation to the number of cycles. A correlation of the security defense framework examination findings is illustrated in Figure 5.

Table 1: Security defense system results in the comparison analysis.

Business Continuity (BC), Access Control (AC), Common Criteria (CC), Corporate Governance (CG)

	BC/AC	**CC/BC**	**CG/CC**
User behavior	30%	59%	15%
SQL Injection	44%	32%	27%
Spam	55%	31%	17%
Access Address	14%	31%	58%

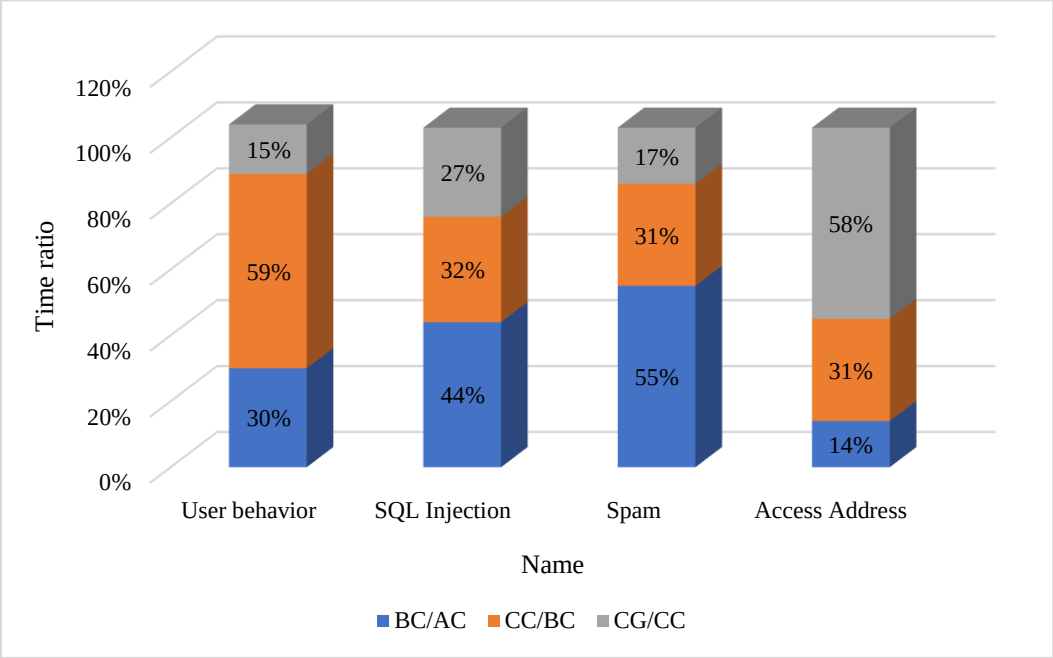


Figure 5: Security defence system findings in the comparative analysis.

A comprehensive overview of the virtual machine preliminary stage of the cautious structure's improvement is provided. The virtual machine's configuration and the associated writing computer programs are thoroughly examined to determine the anticipated reasonable environment for the preliminary stages of the distinctive protective system. The unique defense framework's strength and reasonableness were evaluated in a LAN environment that had been developed. The trial results revealed that the network hosts of one or two types of assigned exercises that were protected by the unique defense framework were generally able to transmit without feeling immediately alert and with minimal impact on the underlying network structure's performance. When an aggressor examines the network protected by the distinctive defense design and endeavors to determine its geographic location, they are unable to view the actual IP address information of the LAN. In fact, the extraordinary defense framework can thwart the aggressor's sifting checks, thereby increasing the attack cost

5. Conclusion and future scope

Quantitative evaluation research and cybersecurity are the topics that are discussed in this article. Following this, the current state of research in the areas of cybersecurity modeling and quantitative assessment is revealed. Lastly, it evaluates the degree to which there are gaps in the research. Comprehensive findings have been obtained through the application of optimization techniques, the measurement of network security defense systems, and the modeling of information system network security. These findings are supported by the fact that the proposed solutions have been demonstrated to be effective in practice. A rule-based algorithm that systematically investigates security device deployment configurations and enables configuration modification is proposed in this article. The goal of the algorithm is to improve and optimize network security defense systems. For the purpose of determining the significance of security devices and differentiated optimization, this foundation is utilized. For the purpose of determining the most effective approach to putting in place a network security defense system, an optimization solution is developed. The implementation of a measurement scheme is a solution to the problem of redundant stacking of security device functions as well as inefficient deployment. For the purpose of developing an effective defense system measurement and optimization system, the proposed network security defense system measurement and optimization scheme is utilized. The system is comprised of modules that are responsible for data entry, defense system evaluation, traversal of deployment methods, and optimization. Defense system evaluation and optimization are both within the realm of possibility. The proposed method for measuring and optimizing the network security defense system in applications that are actual in the real world. The effectiveness of the network security defense system scheme that is described in this article has been demonstrated through the results of optimization and measurement.

When it comes to our future work, one of the most important components will be the enhancement of the framework for anomaly correction. In order to meet the ever-increasing demand for real-time data processing, the framework will be modified to include real-time monitoring of the quality of the data. This includes the investigation of methods for correcting anomalies and monitoring the quality of data in real time.

References

[1] Zarei, S.M.; Fotohi, R. Defense against flooding attacks using probabilistic thresholds in the internet of things ecosystem. Secur. Priv. 2021, 4, e152.

[2] Zhou, Z.; Tian, Y.; Xiong, J.; Ma, J.; Peng, C. Blockchain-enabled Secure and Trusted Federated Data Sharing in IIoT. IEEE Trans. Ind. Inform. 2022; Early Access.

[3] Prvan, M.; Ožegović, J. Methods in Teaching Computer Networks: A Literature Review. ACM Trans. Comput. Educ. 2020, 20, 1–35.

[4] Nour, B.; Mastorakis, S.; Ullah, R.; Stergiou, N. Information-Centric Networking in Wireless Environments: Security Risks and Challenges. IEEE Wirel. Commun. 2021, 28, 121–127

[5] A. Ju, Y. Guo, and T. Li, "MCKC: a modified cyber kill chain model for cognitive APTs analysis within Enterprise multimedia network," Multimedia Tools and Applications, vol. 79, no. 39-40, pp. 29923–29949, 2020.

[6] Bhat, S. A., & Huang, N. F. (2021). Big data and ai revolution in precision agriculture: Survey and challenges. Ieee Access, 9, 110209-110222.

[7] C. Zhou, H. Wang, C. Wang et al., "Geoscience knowledge graph in the big data era," Science China Earth Sciences, vol. 64, no. 7, pp. 1105–1114, 2021.

[8] D. Wang, J. Yu, B. Liu, C. Long, P. Chen, and Z. Chong, "Integrated energy efficiency evaluation of a multi-source multi-load desalination micro-energy network," Global Energy Interconnection, vol. 3, no. 2, pp. 128–139, 2020.

[9] R. Vinayakumar, K. P. Soman, P. Poornachandran, "Applying convolutional neural network for network intrusion detection," In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2017, September, pp. 1222–1228.

[10] H. Sun, Z. Yao, and Q. Miao, "Design of macroeconomic growth prediction algorithm based on data mining," Mob. Inf. Syst., vol. 2021, no. 7, pp. 1–8, 2021.

[11] Jeyepalan, D. P., & Kirubakaran, E. (2019). High performance network intrusion detection model using graph databases. International journal of computational intelligence and information security December 2019.

[12] Essid, M., & Jemili, F. (2016). Combining intrusion detection datasets using MapReduce. proceedings of the International Conference on Systems, Man, and Cybernetics.

[13] Ben Fekih, R., & Jemili, F. (2018). Distributed architecture of an Intrusion detection system based on cloud computing and big data techniques. 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT 2018).

[14] Radhakrishna, V., Aljawarneh, S., Kumar, P. V., Janaki, V., & Cheruvu, A. (2019). Tree based data fusion approach for mining temporal patterns. Proceedings of the 5th international conference on engineering and MIS (ICEMIS '19). association for computing machinery, New York, NY, USA.

[15] Singh, O. P., Singh, A. K., & Zhou, H. (2022). Multimodal fusion-based image hiding algorithm for secure healthcare system. IEEE Intelligent Systems.

[16] Fadhil, S. A., Lubna, E. K., &Sayl, G. A. (2021). Protection measurements of computer network information security for big data. Journal of Discrete Mathematical Sciences and Cryptography, 24(7), 1959–1965. doi:10. 1080/09720529.2021.1959996 doi:10.1080/09720529.2021.1959996

[17] Gupta, R., Tanwar, S., Tyagi, S., & Kumar, N. (2020). Machine learning models for secure data analytics: A taxonomy and threat model. Computer Communications, 153, 406-440.

[18] Hajjaji, Y., Boulila, W., Farah, I. R., Romdhani, I., & Hussain, A. (2021). Big data and IoT-based applications in smart environments: A systematic review. Computer Science Review, 39, 100318.

[19] Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2020). Big data analytics: Computational intelligence techniques and application areas. Technological Forecasting and Social Change, 153, 119253.

[20] J. Hu, S. Cai, T. Huang et al., "Vehicle travel destination prediction method based on multi-source data," Automotive Innovation, vol. 4, no. 3, pp. 315–327, 2021.

[21] Khang, A., Abdullayev, V., Ali, R. N., Bali, S. Y., Mammadaga, G. M., & Hafiz, M. K. (2024). Using Big Data to Solve Problems in the Field of Medicine. In Computer Vision and AI-Integrated IoT Technologies in the Medical Ecosystem (pp. 407-418). CRC Press.

[22] L. Wang, M. Peng, and Q. Zhou, "Pre-impact fall detection based on multi-source CNN ensemble," IEEE Sensors Journal, vol. 20, no. 10, pp. 5442–5451, 2020.

[23] M. S. Tehrany, S. Jones, F. Shabani, F. Mart´ınez-Alvarez, and ´ D. Tien Bui, "A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility

using Logit Boost machine learning classifier and multi-source geospatial data," 5eoretical and Applied Climatology, vol. 137, no. 1-2, pp. 637–653, 2019.

[24] N. Sanden and G. Neideck, "Learnings from the development of public sector multi-source enduring linked data assets," Australian Journal of Social Issues, vol. 56, no. 2, pp. 288–300, 2021.

[25] Oluwafunmilayo G . An Assessment of Cybersecurity Technologies in the Selected Universities in Southwestern Nigeria[J]. International Journal of Computer Applications; 2019; 178(50):11-18.

[26] Pramanik, S., & Bandyopadhyay, S. K. (2023). Analysis of big data. In Encyclopedia of data science and machine learning (pp. 97-115). IGI Global.

[27] Anjum, N., Latif, Z., Lee, C., Shoukat, I. A., & Iqbal, U. (2021). MIND: A multi-source data fusion schemefor intrusion detection in networks.Sensors,21(14), 494.

[28] Aleroud, A., & Karabatis, G. (2017). Contextual information fusion for intrusion detection: a surveyand taxonomy.Knowledge and Information Systems,52(3), 563–619.

[29] Q. Guo, S. Jin, M. Li et al., "Application of deep learning in ecological resource research: theories, methods, and challenges," Science China Earth Sciences, vol. 63, no. 10, pp. 1457–1474, 2020.

[30] Ragazou, K., Passas, I., Garefalakis, A., Galariotis, E., &Zopounidis, C. (2023). Big data analytics applications in information management driving operational efficiencies and decision-making: mapping the field of knowledge with bibliometric analysis using R. Big Data and Cognitive Computing, 7(1), 13.

[31] S. S. Harsha, H. Simhadri, and K. Raghu, "Distinctly trained multi-source CNN for multi camera based vehicle tracking system," International Journal of Recent Technology and Engineering, vol. 8, no. 2, pp. 624–634, 2019.

[32] Tao, D., Yang, P., & Feng, H. (2020). Utilization of text mining as a big data analysis tool for food science and nutrition. Comprehensive reviews in food science and food safety, 19(2), 875-894.

[33] Ullah, F., Salam, A., Abrar, M., & Amin, F. (2023). Brain tumor segmentation using a patch-based convolutional neural network: A big data analysis approach. Mathematics, 11(7), 1635.

[34] Widad Elouataoui, Saida El Mendili, and Youssef Gahi, "An Automated Big Data Quality Anomaly Correction Framework Using Predictive Analysis" Journal Data, MDPI, December 2023.

[35] Gahi, Y.; El Alaoui, I. A Secure Multi-User Database-as-a-Service Approach for Cloud Computing Privacy. Procedia Comput. Sci. 2019, 160, 811–818.

[36] Vasa, J., & Thakkar, A. (2023). Deep learning: Differential privacy preservation in the era of big data. Journal of Computer Information Systems, 63(3), 608-631.

[37] Wu, J., Wang, J., Nicholas, S., Maitland, E., & Fan, Q. (2020). Application of big data technology for COVID-19 prevention and control in China: lessons and recommendations. Journal of medical Internet research, 22(10), e21980.

[38] X. Chen, D. Zhao, W. Zhong, and Y. Jiufeng, "Research on information sharing technology of mental health alliance based on multi-source heterogeneous data fusion algorithms," Academic Journal of Computing & Information Science, vol. 2, no. 1, pp. 74–80, 2019.

[39] Y. Xiong and F. Zhang, "Effect of human settlements on urban thermal environment and factor analysis based on multisource data: a case study of Changsha city," Journal of Geographical Sciences, vol. 31, no. 6, pp. 819–838, 2021.

[40] Z. Xiong, H. Xu, W. Li, and Z. Cai, "Multi-source adversarial sample attack on autonomous vehicles," IEEE Transactions on Vehicular Technology, vol. 70, no. 3, pp. 2822–2835, 2021.