
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 2, pp. 691–698, DOI 10.12694/scpe.v26i2.3962

RESEARCH AND IMPLEMENTATION OF CAMPUS NETWORK INTRUSION DETECTION SYSTEM BASED ON DATA MINING AND IMAGE PROCESSING

ZHE ZHANG*

Abstract. In order to solve the problem of traditional intrusion detection system programs usually being manually written, with a large workload and certain limitations, the author proposes the research and implementation of a campus network intrusion detection system based on data mining and image processing. Its hardware components include a data warehouse, sensors, checkers, generators, etc; The software design includes a packet capture module, a data preprocessing module, and an event analyzer module. Experimental comparison with traditional methods. The experimental results show that the campus network intrusion detection system designed by the author is far superior to traditional system design in detecting intrusion behavior, approaching 100% infinitely, and has high effectiveness. The system has a certain degree of adaptive ability and can effectively detect external intrusions.

Key words: Image processing, Data mining, Campus network, Intrusion detection

1. Introduction. In recent years, with the rapid development of image acquisition and storage technology, we have been able to easily obtain a large amount of useful image data (such as remote sensing image data, medical image data, etc.). But how to fully utilize these image data for analysis and extract useful information from them has become the biggest problem we face. Image data mining has emerged as an emerging field in data mining [1,2]. Image data mining is a technology used to mine hidden knowledge, relationships within or between images, and other patterns hidden in image data on a large scale. It is still in the experimental research stage and is an emerging but highly promising research field. Introducing data mining techniques into image processing is not something that can be achieved overnight. Early image data mining was only focused on certain preprocessing of images, such as image segmentation based on data mining, image feature extraction based on data mining, and so on. With the rapid development of image processing and data mining technology, the early exploratory application of data mining technology in image processing can no longer meet the needs of practical applications. Therefore, the author proposes the idea of further integrating the two disciplines. At this stage, the image is first preprocessed, striving to use a unified representation model and method to process the image data; Then, data mining is carried out on the processed image data to more effectively extract relevant useful data. Among them, the latter stage is currently the focus of research and development in image data mining technology [3,5].

As the Internet continues to expand rapidly, it has become an integral part of daily life, bringing with it a myriad of challenges to network security. The persistent threat of hackers infiltrating computer networks and the onslaught of virus attacks have presented significant hurdles for networks and information systems. In response, diverse measures have been implemented to safeguard network systems. These include digital signature technology, biometric technology, content filtering technology, network isolation technology, and more. While these defense mechanisms offer a degree of protection against external interference and ensure swift and efficient information transmission, they also suffer from drawbacks such as high rates of false positives and false negatives[6]. Consequently, accurately and effectively detecting attacks remains a formidable challenge. Traditional intrusion detection systems are constrained by their ability to only detect data from established databases. As the frequency and complexity of attacks escalate, these databases expand, leading to decreased detection efficiency and system overload. In contrast, network intrusion detection utilizing data mining techniques delves into vast datasets to uncover hidden and previously unknown valuable insights. By accurately

*School of Computer Science and Technology, Nanyang Normal University, Nanyang, Henan, 473061, China (Corresponding author, ZheZhang56@126.com)

identifying correlations among intrusion data, it sidesteps the need for labor-intensive manual analysis and minimizes the associated workload and human intervention. This approach ultimately curtails operational and maintenance expenses [7].

2. Literature Review. Image data mining refers to extracting or mining useful information or knowledge from large-scale image sets. Therefore, we can understand image data mining as an application of data mining in the field of images [8]. The two fundamental concepts of image data mining are "large-scale image sets" and "extracting and mining useful information and knowledge.". From the perspective of "large-scale image sets", it involves fields such as image acquisition, image storage, image compression, and multimedia databases; From the perspective of "mining useful information and knowledge", this approach encompasses a range of disciplines including image processing and analysis, pattern recognition, computer vision, image retrieval, machine learning, artificial intelligence, and knowledge representation. Therefore, image data mining is a multidisciplinary emerging field, and most of the other fields it involves are also in the development stage, and it itself is also in the experimental stage. Intrusion detection, as a key component of computer network security, involves multiple technologies such as daily log parsing, system vulnerability checking, and network link detection. Intrusion detection systems can be categorized into two main types: anomaly-based detection technology and misuse-based detection technology. Anomaly-based detection technology assesses current system user actions against typical user behavior to identify any deviations indicative of intrusion activity [9-10]. Brahma, A. et al. devised a collaborative neural fuzzy inference system that integrates genetic algorithm-based database intrusion detection systems. This innovative approach proves adept at identifying malicious transactions within databases. Through rigorous experimental analyses and comparative assessments against established statistical database intrusion techniques, the efficacy of the proposed system was convincingly demonstrated [11]. Attou, H. et al. introduced a cloud-based intrusion detection model leveraging random forest and feature engineering. The model incorporates a random forest classifier, enhancing the accuracy of the detection system [12]. Wang et al. introduced an innovative intrusion detection approach based on deep learning principles. Their method utilizes two separate in-memory autoencoders trained on both normal network traffic and attack patterns. This allows for the capture of dynamic relationships between traffic features, particularly in the context of imbalanced training data. Subsequently, the original data is inputted into a triplet network for training, with triplets formed by the reconstructed data from the two encoders. Ultimately, the distance relationship within these triplets serves as the basis for determining whether a given traffic instance constitutes an attack [13-14].

With the gradual emergence of various security threats, it also drives continuous innovation in information network security. A large number of information security protection technologies have emerged, including host security protection technology, data encryption technology, network firewall, website security access technology, identity authentication technology, and vulnerability scanning technology. Due to the widespread potential information security issues in computer networks, the above technologies are unable to detect and respond to intrusion attacks in a timely manner, resulting in serious economic losses. Therefore, it is particularly important to construct a proactive and proactive security protection system that can actively prevent attacks and intrusion behaviors. Unlike passive defense firewall technology, as an active attack security protection technology, it can monitor network data in real-time, detect various attacks or abnormal behaviors inside and outside the network, and actively respond and block attacks before they occur or cause serious economic losses. Intrusion detection systems are a good choice for addressing popular information security threats such as network attacks and cross site scripting attacks.

The author mainly designs a high-level template library for a low code management system based on the results of system requirements analysis. Firstly, the overall architecture of the system is designed according to the overall requirements of the system. Then, based on the results of functional requirements analysis, the system functional modules are designed, and the user roles and permissions existing in the system are divided. Finally, the system directory structure is designed.

3. Method.

3.1. Intrusion detection technology and classification. In the 1980s, intrusion detection technology (IDS) emerged as a proactive defense mechanism. By gathering crucial data from computer systems and subjecting it to analysis, IDS can identify potential security policy breaches and network attacks. Intrusion

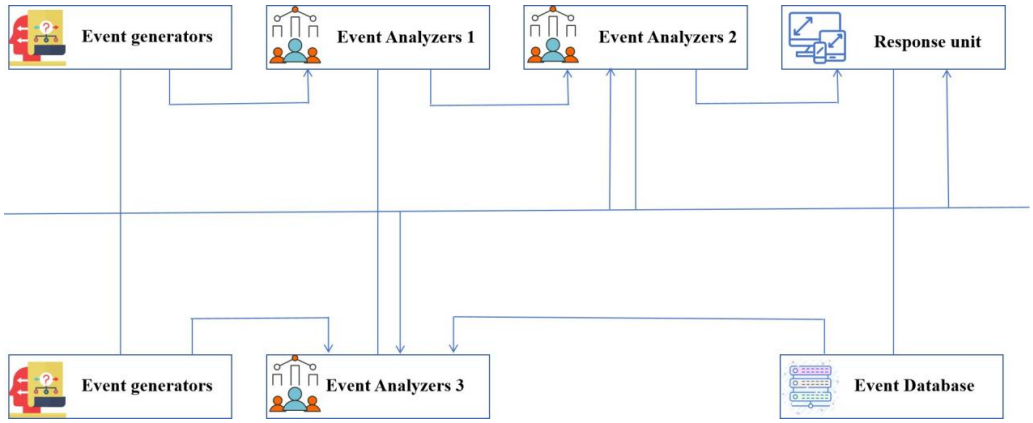

Fig. 3.1: Intrusion Detection System Model

behavior typically falls into two categories: internal and external. Internal intrusion pertains to unauthorized actions by legitimate users, whereas external intrusion involves the infiltration of hackers or unauthorized individuals. Such intrusive activities pose risks to the integrity, functionality, and confidentiality of network data, and various intrusion detection technologies prioritize different aspects of this threat landscape. Intrusion detection can be categorized based on the outcomes they aim to achieve, resulting in anomaly detection and misuse detection, which employ distinct detection methods[15]. Additionally, they can be classified as network-based or host-based, depending on the data sources they utilize. Furthermore, intrusion detection systems can be distinguished by their detection timing, either real-time or non-real-time. Models such as the unified model and the Snort model are prevalent in this domain, typically comprising event generators, analyzers, response units, and databases, as illustrated in Figure 3.1.

3.2. Process Analysis of Intrusion Detection. When dealing with intrusion behavior, the detection of network systems typically involves four key stages:

1. Data Collection: This initial phase involves gathering information about the system through various means such as external sensors or proxy hosts. It includes capturing user behaviors, system statuses, and fundamental network data.
2. Data Processing: In this stage, the collected data undergoes processing and transformation into a standardized format recognized by computers. This step aims to enhance the efficiency and speed of detection.
3. Data Analysis: Here, the processed information is analyzed. This may involve comparing patterns with known databases, conducting statistical analyses using probability theory, and identifying potential threats. Uncertain or suspicious data is forwarded to the control module for further assessment.
4. System Response: Based on the analysis and in accordance with predefined rules, the system responds accordingly. This could involve actions such as reconfiguring routers, isolating intruder IPs, or modifying file properties to mitigate the intrusion.

This process is illustrated in Figure 3.2.

3.3. Data mining techniques. Data mining involves uncovering potentially significant patterns or rules within vast amounts of complex, noisy, and inconsistent data. The implementation of data mining typically encompasses three primary processes: (1) Data Preparation: This initial stage involves various tasks such as selecting appropriate data targets, identifying operational entities, preprocessing data to remove noise, and reducing the dimensionality of the data. These steps aim to make the data suitable for analysis. (2) Data Mining: In this phase, utilizing different data mining models, suitable mining algorithms are chosen to extract valuable insights from large, incomplete, and irregular datasets. The goal is to uncover patterns or relationships that can help predict outcomes or discover hidden knowledge. (3) Data Representation and Evaluation:

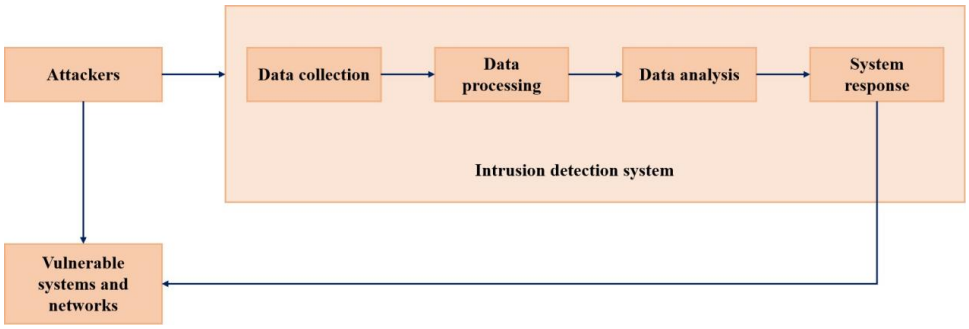

Fig. 3.2: Intrusion Detection Process

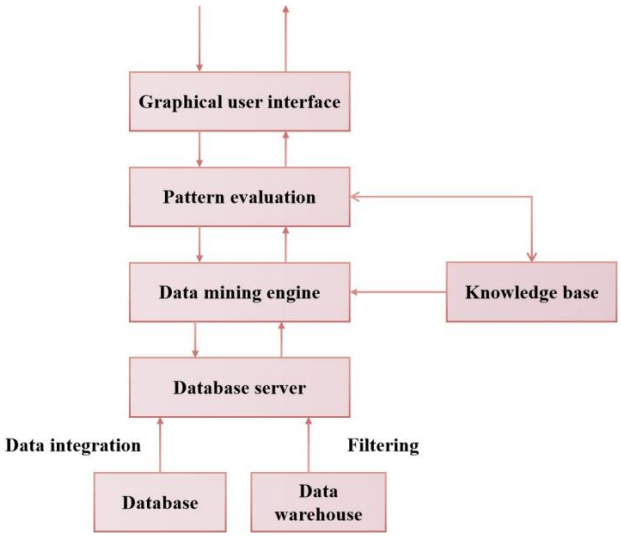

Fig. 3.3: Intrusion Detection Process

Following data mining, the acquired information undergoes analysis through techniques such as association rule mining, classification, and clustering. This process aims to derive meaningful insights and present them in a comprehensible manner, often through data visualization techniques. The structure of a data mining system is depicted in Figure 3.3.

3.4. Data Mining Algorithms for Intrusion Detection. Intrusion detection systems heavily rely on data mining algorithms for effective threat detection. Various algorithms come with distinct pros and cons, suitable for different models. Statistical analysis, feature analysis, change and deviation analysis, and clustering are frequently employed methods in data mining. Among these, association rules play a central role, revealing the connections within data. Mathematically, association rules can be defined as follows: Suppose there's a database D containing m pieces of information, denoted as Ti, where each piece of information comprises n units I. The relationship between these units is depicted as $D = \{T_1, T_2, \cdots, T_m\}, T = \{I_1, I_2, \cdots, I_n\}$, the subset of the database is represented by A. If there is | A |=K, then A is called the project set of K. The support of database D containing project set A is represented by Cx. If there are project sets A and B, then their association rules are:

$$Support(A \Rightarrow B) = P(A \cup B) \tag{3.1}$$

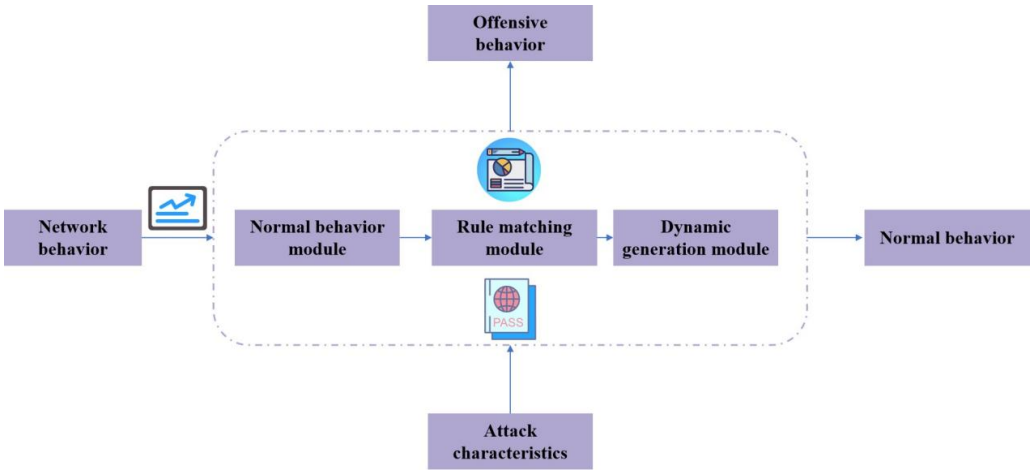

Fig. 3.4: Intrusion Detection Process

$$Confidence(A \Rightarrow B) = P(A|B) \tag{3.2}$$

In equations 3.1 and 3.2, Support(A) signifies the frequency of occurrence of itemset A within the database D, while Confidence(A) denotes the strength of the association rule for A. The calculation method:

$$Support(A)/\% = \frac{C_x}{|D|} \times 100 \tag{3.3}$$

Hence, by examining the support and confidence values between project sets A and B, we can ascertain the association rules between them. If the confidence of association rules A and B exceeds the minimum confidence threshold for the project set and the support of A and B is equal to or greater than the minimum support threshold, it indicates strong association rules between the project sets A and B. Conversely, if these conditions are not met, it suggests weak association rules between them. We can also gauge the correlation between project sets A and B through another metric:

$$r_{A,B} = \frac{\sum(A - \overline{A})(B - \overline{B})}{(n-1)\sigma_A \sigma_B} \tag{3.4}$$

In equation 3.4, the average values of A and B are $\overline{A}$, $\overline{B}$, and the standard deviation is $\sigma_A \sigma_B$.

3.5. Overall Design of Intrusion Detection System. To ensure timely and effective analysis of network data, the author emphasizes the essence of intrusion detection system design: mining association rules and sequence rules between data, and implementing classification recognition based on these rule definitions. Since different system models require different data mining algorithms, selecting a suitable intrusion detection system becomes crucial. For this purpose, the author suggests using the lightweight open-source intrusion detection system, Snort. Despite its effectiveness against most network attacks, Snort suffers from limitations such as low efficiency, false alarms, missed reports, and the inability to perform dynamic real-time detection. Consequently, there's a need to enhance the Snort detection model[16]. The enhanced Snort system model is illustrated in Figure 3.4.

3.6. Hardware composition of campus network intrusion detection system based on data mining. With the comprehensive analysis of hardware requirements for intrusion detection systems, the author designed the hardware part of a campus network intrusion detection system based on frequency hopping data mining. This part mainly consists of a data warehouse, sensors, inspectors, generators, etc. Its advantages are stable performance, good scalability, and the ability of multiple intrusion detection components in the

network to work together. Sensors can timely capture raw data appearing in computer networks, obtain basic characteristics of system data analysis, and upload formatted data to inspectors. The inspector uses a detection model to detect the data structure of the sensor uploaded data, detect whether there is intrusion behavior, and upload the results to the data warehouse. As the application storage center of the entire system, a data warehouse is mainly used to store historical data information. A generator refers to the ability to process detected abnormal data or intrusion behavior in a timely manner, generate a special type of data, and then connect this data with the historical information stored in the data warehouse to obtain new intrusion features, and put them into the intrusion feature storage library, which is conducive to early detection when encountering the same attack. When the system detects abnormal data, the alarm will alert [17].

3.7. Software composition of campus network intrusion detection system based on data mining.

3.7.1. Packet Capture Module. Wincap is mainly composed of three parts, namely the packet filtering driver Packet.dll, the lower level data dynamic connection library, and API that can directly access driver data. Wincap can capture a large number of data packets on computer networks and filter them according to pre-set patterns. The execution steps for capturing data packets using Wincap are as follows. Firstly, obtain the adapter sequence table and the model of the network adapter in the system. Secondly, select a default adapter number 1 from all adapters and set it to mixed mode. Once again, set the packet filtering driver in BPO to filter the original data packets, such as interfaces, IP addresses, etc. The buffer pool needs to have 564 K of storage space, allocate a packet object, and link to the allocated buffer pool; Assign packet matching objects and link them to a pre allocated buffer pool to capture multiple driver packets; And receive the target data packet from the network Lap Adaptation, and put the data packet into the packet structure targeted by the Lap Adaptation. Once received, it immediately returns True, otherwise it returns FALSE. Finally, with the help of a triggering device and a mathematical function model, the captured data packets that comply with the filter rules are uploaded to the network protocol analysis module for analysis. After completion, the processing object of the data packet is released, and the network card instrument is shut down to restore the network card to its normal state.

3.7.2. Data preprocessing module. Data processing refers to the large-scale cleaning, collection, transformation, discretization, and classification of raw data, providing clear and reasonable data preparation for data mining. Data processing refers to the pre-processing of captured data packets for subsequent data matching and intrusion detection processes. The preprocessing module mainly serves BPO filtering and decoding of network protocol code. Protocol decoding refers to the programmatic processing of data packets. The captured raw network packets cannot be directly applied and can only be used after being manipulated. This is because network packets are divided into several data slices with their own feature codes before uploading, but Internet networks are based on group conversion programs, so the order and sequence of uploading this data slice to the host is irregular, and the data uploaded to the host system is an unordered data stream. In order to effectively reply to the data stream, it is necessary to recombine the data stream. Therefore, the processing methods vary for different network protocols [18,19].

3.7.3. Event Analyzer Module. By processing the aforementioned packet capture module, the original network packets were obtained, but data mining cannot be directly used to apply them in intrusion detection systems. Before application, the initialization data packet must be processed according to data mining methods to extract main features. After data preprocessing, it must be checked according to the classifier of the data structure. The event analysis module mainly includes two processing stages, the training stage and the actual testing simulation stage. During the training phase, a pre selected combination of data is used, followed by a predetermined plan. After extracting the main features and preprocessing the data, the best classification results can be obtained through training; After capturing the basic data packets in the actual running network during the testing simulation phase, the same feature extraction and data preprocessing operations will be performed on them, and then the classifier obtained during the training phase will be used to monitor the system for intrusion behavior [20].

Table 3.1: Test Environment Configuration Information

Project	Configuration information
Database CPU	Intel(R)Core(TM)i7 @2.26 GHz 2.26 GHz
Business processing server memory	DDR38 G
Monitoring center bandwidth	1000 M/s
Switch model	Cisco 3350
storage	Flash 3G

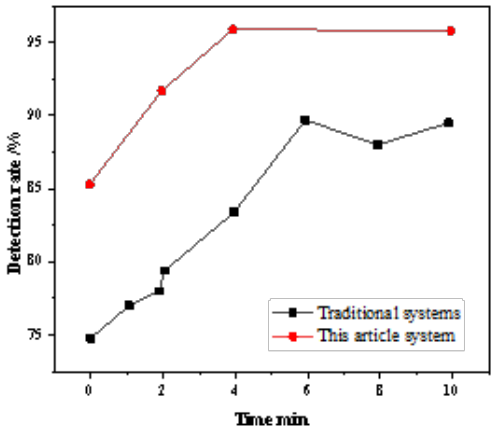

Fig. 4.1: Comparison of experimental results

3.8. Experimental Preparation. In order to more clearly and specifically demonstrate the practical application effect of the campus network intrusion detection system designed by the author based on frequency hopping data mining, a comparison is made with traditional campus network intrusion detection systems to compare their detection rate capabilities. In order to ensure the accuracy of the experiment, two campus network intrusion detection system designs were placed in the same testing environment for detection rate capability experiments. The test environment configuration is shown in Table 3.1.

4. Results and Discussion. During the experiment, two different campus network intrusion detection systems were designed to work simultaneously in the same environment, and the changes in detection rate and capability were analyzed. The comparison of experimental results is shown in Figure 4.1.

The experimental results show that the campus network intrusion detection system designed by the author based on frequency hopping data mining is far superior to traditional system design in detecting intrusion behavior, approaching 100% infinitely, and has high effectiveness.

5. Conclusion. The author advocates for developing and deploying a campus network intrusion detection system that leverages both data mining and image processing techniques. The rapid development of the network not only enriches people's lives, but also brings many network security issues. As an important part of network protection, network intrusion detection systems play a very important role. As the network scale grows and the variety of network attacks increases, conventional intrusion detection systems are struggling to effectively adapt to the evolving environment. The author first analyzed the classification and models of intrusion detection, as well as commonly used data mining algorithms. By integrating data mining techniques into intrusion detection, the author conducted comprehensive data analysis and successfully developed a detection system. This system's accuracy and detection speed were evaluated against other existing systems, determining its

practical significance. At the same time, considering the multitude of intrusion behaviors, the future focus of intrusion detection technologies lies in enhancing application layer intrusion detection and implementing adaptive intrusion detection mechanisms.

REFERENCES

[1] Ekpotu, W. F. , Akintola, J. , Obialor, M. C. , & Philemon, U. . (2023). Historical review of hydrogen energy storage technology. World Engineering and Technology (in English), 11(3), 454-475.
[2] Kim, E. , An, J. , Cho, H. C. , Cho, S. , & Lee, B. . (2023). A sensor data mining process for identifying root causes associated with low yield in semiconductor manufacturing. Data Technologies and Applications, 57(3), 397-417.
[3] Qu, Y. , Samarati, P. , Benslimane, A. , & Yu, S. . (2022). Call for papers special issue on privacy-preserving data mining for artificial intelligence of things. Big Data Mining and Analysis, 5(1), 1.
[4] Zhi-wenZHANG, Tian-geLIU, & Peng-juNIE. (2022). Real-time semantic segmentation for road scene based on data enhancement and dual-path fusion network. Acta Electronica Sinica, 50(07), 1609-1620.
[5] Saesi, N. , & Taleghani, M. . (2023). Linking competitors' knowledge and developing innovative products using data mining techniques. Computer and Communication (English), 11(7), 37-57.
[6] Xue-songWANG, Han-linZHANG, & Yu-huCHENG. (2022). Autoencoder and hypergraph-based semi-supervised broad learning system. Acta Electronica Sinica, 50(03), 533-539.
[7] Li, Q. Q. , Wang, G. K. , Liang, Z. X. , & Hu, Z. J. . (2022). Highly transparent and adhesive poly(vinylidene difluoride) films for self-powered piezoelectric touch sensors. Chinese Journal of Polymer Science, 40(7), 726-737.
[8] Kumar, A. , Abhishek, K. , Ghalib, M. R. , Shankar, A. , & Cheng, X. . (2022). Intrusion detection and prevention system for an iot environment. Digital Communication and Networking: English Version, 8(4), 540-551.
[9] Al-Kahtani, M. S. , Mehmood, Z. , Sadad, T. , Zada, I. , Ali, G. , & Elaffendi, M. . (2023). Intrusion detection in the internet of things using fusion of gru-lstm deep learning model. Intelligent Automation and Soft Computing, 37(8), 2279-2290.
[10] Cui, L. . (2022). Information security management measures for college archives under the network environment. Electronic Research and Applications, 6(6), 15-19.
[11] Attou, H. , Guezzaz, A. , Benkirane, S. , Azrour, M. , & Farhaoui, Y. . (2023). Cloud-based intrusion detection approach using machine learning techniques. Big Data Mining and Analytics, 6(3), 311-320.
[12] Brahma, A. , Panigrahi, S. , Samal, N. , & Gountia, D. . (2022). Rule-based database intrusion detection using coactive artificial neuro-fuzzy inference system and genetic algorithm. International journal of business intelligence and data mining, 6(1), 32-43.
[13] Attou, H. , Guezzaz, A. , Benkirane, S. , Azrour, M. , & Farhaoui, Y. . (2023). Cloud-based intrusion detection approach using machine learning techniques. Big Data Mining and Analytics, 6(3), 311-320.
[14] Wang, W. , Li, J. , Zhao, N. , & Liu, M. . (2023). Mem-tet: improved triplet network for intrusion detection system. Computer, material, and continuum (English) (7), 471-487.
[15] Akanni, A. , Akanni, W. , & Daso, O. H. . (2023). Captcha-based honey net model against malicious codes. Computer and Communication (English), 11(3), 159-166.
[16] Ding, S. , Kou, L. , & Wu, T. . (2022). A gan-based intrusion detection model for 5g enabled future metaverse. Mobile Networks and Applications, 27(6), 2596-2610.
[17] Alsemairi, S. S. . (2023). Factors influencing employees on compliance with cybersecurity policies and their implications for protection of information and technology assets in saudi arabia. Intelligent Information Management, 15(4), 259-283.
[18] Wang, C. , Chen, R. L. , & Gu, L. . (2023). Improving performance of virtual machine covert timing channel through optimized run-length encoding. Journal of Computer Science and Technology, 38(4), 793-806.
[19] Nagamunthala, M. , & Manjula, R. . (2023). Implementation of a hybrid triple-data encryption standard and blowfish algorithms for enhancing image security in cloud environment. Computer and Communication (English), 11(10), 135-149.
[20] Gihonia, S. A. , Mabela, R. M. , René Gilles Bokolo, Kimba, E. , Katshitshi, M. , & Kalombo, M. , et al. (2022). Intrusion detection system with remote signalling for vehicles using an arduino controller and radio-frequency technology. Software Engineering and Applications (English), 15(4), 14.
[21] Sezgin, A. , & Boyac, A. . (2023). Aid4i:an intrusion detection framework for industrial internet of things using automated machine learning. Computers, Materials, and Continuum (in English), 76(8), 2121-2143.

Edited by: Bradha Madhavan
Special issue on: High-performance Computing Algorithms for Material Sciences
Received: May 17, 2024
Accepted: Jun 25, 2024