Hindawi
Journal of Computer Networks and Communications
Volume 2022, Article ID 9049420, 10 pages
https://doi.org/10.1155/2022/9049420

Research Article

Optimal Management of Computer Network Security in the Era of Big Data

Minfeng Chen

Wuxi Vocational Institute of Commerce, Wuxi 214100, China

Correspondence should be addressed to Minfeng Chen; chenminfeng@wxic.edu.cn

Received 20 August 2022; Revised 12 September 2022; Accepted 21 September 2022; Published 5 October 2022

Academic Editor: Arvind Kumar

As the "new oil of the future," big data is becoming the leading industry of the new economy, the core asset of the country and enterprises, the "new blue ocean" to be pursued, and the national strategy to be developed by all countries. The development of big data and its related technology supports and promotes a new round of technological innovation, making a new generation of information security technology reform and innovation, bringing opportunities and challenges to optimize, and consolidating national information security. In the era of big data, what kind of challenges and impacts will information security face? and is it crucial to explore the response strategies? At present, China has risen to become the world's largest number of Internet users and the largest number of people using smartphones, but because China's information security is the initial stage, involving information security, especially national information security laws and regulations are not much, the national social supervision and monitoring mechanisms are not much, the application level of science and technology content is relatively backward, the core technology has a patent technology not much, resulting in the flood of network data nowadays. Therefore, the underground illegal "data industry chain" activities are rampant. Therefore, this paper proposes a security-aware model based on the combination of distributed data analysis technology and data features. The model uses data features to dynamically generate a library of situational anomalies, effectively solving the problem of analyzing and processing rapidly and dynamically generated data streams, increasing the detection rate to more than 98%, effectively reducing the possibility of false detection, and having good results on large-scale datasets.

1. Introduction

In the twenty-first century, new technologies such as the Internet, the Internet of things, and cloud computing are developing rapidly, and the amount of data generated by modern society is growing at an unprecedented rate; we are now living in the age of big data [1]. The power of data allows people to make decisions directly without thinking about the causes, which is why big data has caused major disruption in the world [2]. In 2011, a study by the McKinsey Global Institute found that data are gradually becoming an important factor of production and is permeating all McKinsey Global Institute found in 2011 that data are becoming an important factor of production and that big data is permeating all industries and businesses, that the use of big data heralds a new wave of productivity growth and consumer surpluses [3], and that "big data" is "the next frontier for innovation, competition, and productivity improvement" [4].

In the late twentieth century, Hollywood actor King Carey starred in the surrealist film The Truman Show, in which the main character Truman is trapped from birth in a giant studio full of cameras while people around the world watch his reality show[5]. In the age of big data, it has long been unnecessary to monitor your behavior through cameras, the products you searched for on Taobao, the videos you watched on Youku, the food you ordered on Meituan, and so on. Even if you are not connected to the web, the server records all transactions made via computers, tablets, mobile phones or smart wearables. Your shopping habits, your heart rate, or your credit rating become part of big data [6]. Analyzing big data allows us to make many smart decisions: for example, whether razors and nappies sell better together or whether a typhoon boosts sales of vanilla-flavored ice cream, etc., but the security of our data is invisibly compromised [7].

China is currently in a period of rapid development, the external and internal environment is changing rapidly, and the issue of information security is attracting a lot of public attention [8]. In the era of big data, information security in the new era is not only threatened by attacks from cyber hackers but also by low-security awareness among citizens, a backward security education system, etc. The reasons for this are many and include various factors such as the government, society, universities, and the public itself [9]. The increasing number of phone scams, pornographic photo documents, data leaks, portal server crashes, and other incidents in recent years have revealed the low level of security awareness in China and the lack of skills to cope with the consequences of security incidents and highlighted the lack of adequate security training for Chinese university students. Information technology in China started to develop late, and since the beginning of the new century, information technology in China has been developing very rapidly. While the relevant regulations are not stable, the security assurance system is not perfect, and the lack of security awareness and the backwardness of information technology have together become the limiting factors affecting information security in China [10].

Therefore, it is necessary to analyze the current situation of information security education in China, identify the reasons for its insufficiency, and propose countermeasures for improvement in order to improve the level of information security protection in China and ensure the security of personal data, property, privacy, and other aspects of the public. The security and stability of universities and society can only be ensured if the security of public information is guaranteed. Only when the environment is stable and its security is guaranteed can citizens focus on their work and studies and contribute to the country and society.

People have always looked for ways to make science and technology serve people better, and in the age of big data, this proposal that people have fought for has come back to us. While we are reaping the benefits of big data, we are also facing a hitherto unknown impact on information security. With frequent media coverage of information security incidents and increasingly complex threats to information security, systematic and effective information security education is becoming increasingly important. The rapid development of the masses' awareness and skills in information security protection is also an issue that information security education in China needs to address urgently [11].

These are the best of times
but also the worst of times.

2. Related Work

Big data has been a hot research topic in recent years, both at home and abroad, but there is no single definition of big data. According to Wikipedia, big data is information so vast that it cannot be collected, managed, processed and organized into valid information in a short time by using conventional software tools [12]. The McKinsey Global Institute, in its article, "Big Data: the next frontier for innovation, competition and productivity," defines big data as data that are beyond the ability of traditional database software tools to collect, store, manage and analyze [13]. "According to "Steve" (IBM's "digital gatekeeper"), the era of networked big data is characterized by the full integration of the triple world of people, machines and things, unprecedented growth in the amount of data, and the sheer complexity of data patterns—the hallmarks of the era of networked big data [14]. McAfee, the world's largest professional security technology company, views big data as an intelligent activity aimed at turning big data insights into business advantage and a prerequisite for analytics [15]. Schoenberg's definition is more concise and clear: big data is a large amount of data [16].

Cybersecurity is a 5 v characteristic of volume, variety, veracity, velocity and, value [17]. The study [18] of network data is characterized by large scale, burstiness, and suddenness, which makes it difficult for researchers to assess and predict its changing state. The study [19] can only make good use of data if the data flow is controlled. The study [20] proposed a requirements analysis model for data quality that considered many candidate metrics and selected the required metrics according to the needs. The study [21] addressed the issues of how to reduce storage costs, fully utilize computational power, improve throughput, and support distributed nonlinear selection algorithm optimization. The study [22] pointed out that data analysis tends to use fewer fields, so the column storage rate is high, and popular databases in the industry (Bigtable, HBase) are implemented based on column storage. The study [23] proposed a hybrid row-column data storage structure that solves the problem of fast loading, querying, and efficient storage of data. Research [24] develops an advanced, highly scalable, petabyte-scale distributed data storage framework while optimizing the distributed storage structure for data layout to reduce cost and improve efficiency, thus realizing a highly available data distributed storage system. Researchers [25] use computing technology to mine and analyze data, discover the knowledge contained in it, and study the environment, change patterns, and development trends, which are the main ways to explore the deep value of data and realize computable behavior. With the advent of the data era, the complexity and scale of data have grown exponentially, leading to serious bottlenecks in the practicality and performance of traditional mining and computing methods. As a result, data processing techniques have become an important research topic. Divide and conquer strategies are often used to process data, i.e., decompose the data problem into smaller subproblems and combine the solutions of the subproblems to obtain the final solution [26].

In conclusion, data analysis is the core of data processing, and data-based security-awareness research uses data analysis as an effective method for sensing various anomalous behaviors in networks. Currently, there are some problems with security-aware models, such as low accuracy of security-aware results, poor prediction accuracy, coarse

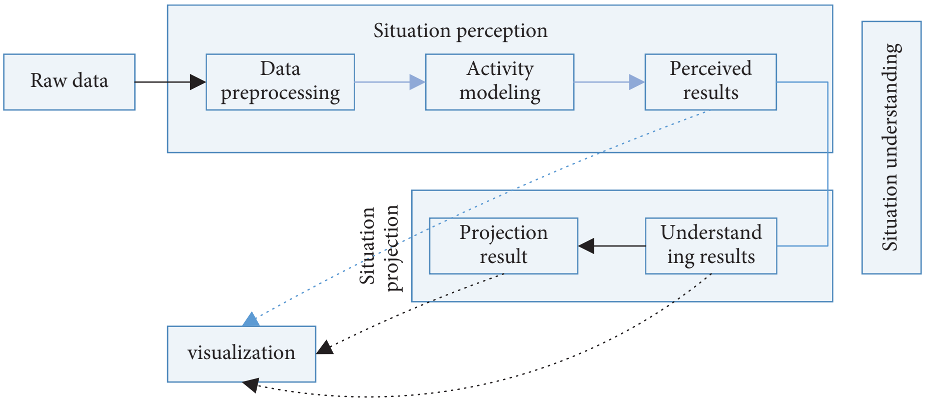

Figure 1: Security situation awareness model.

evaluation granularity, and low perception performance and efficiency. Therefore, our proposed model integrates data preprocessing in perception and performs association rule analysis based on datasets in the situation understanding process. As a result, the performance and efficiency of the model are improved in the security situational awareness process.

3. Network Security Architecture

Network security situational awareness refers to extracting relevant elements in the environment within a certain space and time range, understanding these elements and predicting their possible impacts. Security situational awareness is the cognitive process of system security status. It is generally believed that the first stage is security situation awareness; the second stage is security situation understanding; and the third stage is security situation projection. (Figure 1.

Figure 2 shows that the first stage of situational awareness is carried out at the root node at the top; then, the results of situational awareness are judged to enter the second stage of situational understanding; finally, in the third layer, the situation projection of the third stage is carried out, and the final result of the security situation is obtained.

Figure 3shows that situational awareness and situational understanding are carried out in the peripheral module; then, the results of situational understanding are transmitted to the core module for situational projection.

4. Improved Random Forest Correlation Algorithm

The introduction of decision trees is in the following section, through the decision tree algorithm, to further understand the random forest algorithm.

The calculation formula of entropy is shown in formula (1):

$$E(Y) = \sum_{t=1}^{N} -\frac{Y_t}{\mathrm{Sum}\,(Y)}\log_2\left(\frac{Y_t}{\mathrm{Sum}\,(Y)}\right)\lim_{x\longrightarrow\infty}. \qquad (1)$$

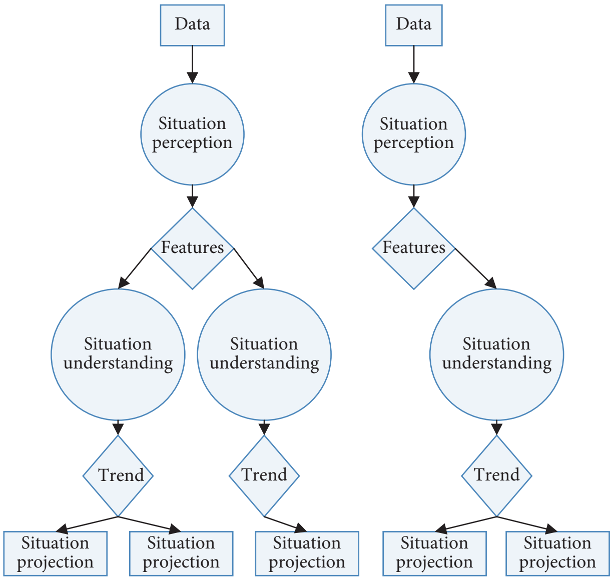

Figure 2: security situation awareness model based on a random forest.

Among them, Y_t represents T rd value of category Y, Sum (Y) represents the total number of records of category Y, and N represents a total of Y values of category N.

Next, we find gain of attribute x. Assuming that attribute X can take a total of M values, value of X_j for one of attributes of x is shown in formula (2):

$$E\left(X_j\right) = \sum_{t=1}^{N} -\frac{Y_t}{\mathrm{Sum}\left(X_j\right)}\log_2\left(\frac{Y_t}{\mathrm{Sum}\left(X_j\right)}\right), \qquad (2)$$

where Sum (X_j) is number of records containing X_j in data set and Y_t is number of records classified as Sum (X_j); gain of attribute Y_t is shown in the following formula (3):

$$E(X) = \sum E\left(X_j\right) = \sum_{j=1}^{M}\sum_{t=1}^{N} -\frac{Y}{\mathrm{Sum}\left(X_j\right)}\log_2\left(\frac{Y}{\mathrm{Sum}\left(X_j\right)}\right). \qquad (3)$$

Bayes' theorem is shown in formula (4):

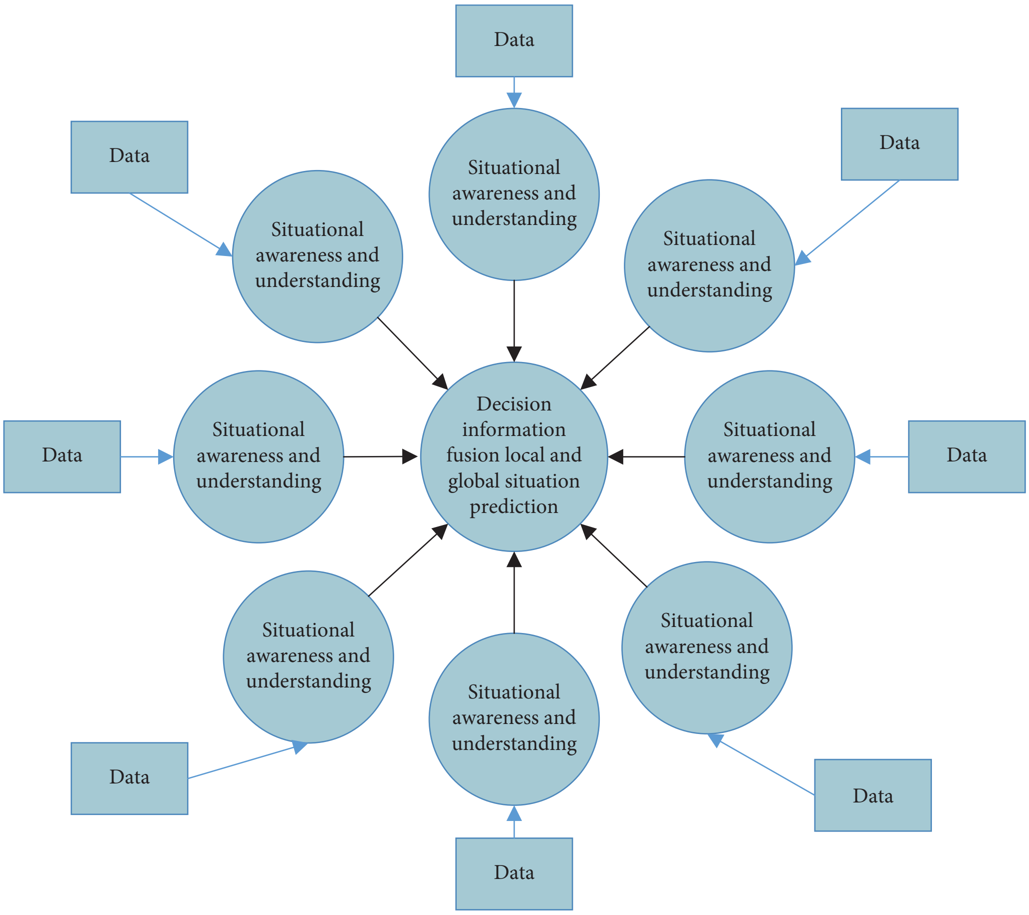
Data
Data
Data
Situational awareness and understanding
Situational awareness and understanding
Situational awareness and understanding
Data
Situational awareness and understanding
Decision information fusion local and global situation prediction
Situational awareness and understanding
Data
Situational awareness and understanding
Situational awareness and understanding
Situational awareness and understanding
Data
Data
Data

Figure 3: security situation awareness model based on a star structure.

$$p(H|x) = \frac{p(x|H)p(H)}{p(x)}. \tag{4}$$

If number of categories in a given dataset is M, the Naive Bayes algorithm can be used to predict whether a given value belongs to category with largest posterior probability, that is, when x is predicted to belong to a certain class C_t by the Naive Bayes classification algorithm, if and only if.

$$p(C_t|x) > p\left(C_j|x\right) 1 \le j \le m, \quad j \ne t. \tag{5}$$

If $p(C_t|x)$ is maximized at this time, then category C_t with largest $p(C_t|x)$ is called maximum a posteriori hypothesis. According to Bayes' theorem,

$$p(C_t|x) = \frac{p(x|C_t)p(C_t)}{p(x)}. \tag{6}$$

For all categories $p(x)$ are equal, and only largest $p(x|C_t)p(C_t)$ can be calculated. When predicting classified samples x of unknown classes, by estimating value of $p(x|C_t)p(C_t)$ corresponding to each class C_t, then sample x belongs to class C_t if and only if.

$$p(x|C_t) > p(C_t)\frac{1s}{s_m}, \quad j = i. \tag{7}$$

Therefore, Bayesian classification methods are mostly used to classify scenes.

The algorithm compresses data into memory in process of building a book, so that the data set only need to be scanned twice, which greatly reduces the overhead of $I/0$, so it has great advantages when dealing with large data.

In order to better describe the whole process, the following symbols are defined, S_{mtn} represents support degree.

$$S = \frac{\operatorname{count}(x \cup y \cup z)}{n}. \tag{8}$$

Formula (8) is the calculation formula of support degree, in which $\operatorname{count}(x \cup y \cup z)$ represents number of data records containing the attribute of x, y, z. Suppose there are 5 data records, each containing the following attributes:

$$\begin{aligned} R_1 &= \{a, c, d, f, g, t, m\}, \\ R_2 &= \{a, b, c, f, l, m, o\}, \\ R_3 &= \{b, f, h, j, o\}, \\ R_4 &= \{b, c, k, s, p\}, \\ R_5 &= \{a, f, c, e, l, p, m, n\}. \end{aligned} \tag{9}$$

Suppose $S_{mtn} = 3$. The process of mining association rules for the aboveentioned data records is as follows:

(1) Scan data for the first time and then generate a 1-dimensional frequent itemset as follows:

$$\begin{aligned} R_1 &= \{c, f, a, m, p\},\\ R_2 &= \{c, f, a, b, m\},\\ R_3 &= \{f, b\},\\ R_4 &= \{c, b, p\},\\ R_5 &= \{c, f, a, m, p\}. \end{aligned} \tag{10}$$

(2) Using 1-dimensional frequent item sets to generate FP-tree, generated complete FP-tree

(3) Mining association rules in generated FP-tree and obtain frequent item sets formed by each attribute that meets minimum support threshold. Since some of generated frequent item sets contain redundant repetitions, simple redundancy removal is performed

The final data record after dimensionality reduction is as follows:

$$\begin{aligned} R_1 &= \{a, c, f, m\},\\ R_2 &= \{a, c, f, m\},\\ R_3 &= \{\varnothing\},\\ R_4 &= \{c, p\},\\ R_5 &= \{a, f, c, m\}. \end{aligned} \tag{11}$$

In order to better describe the data matching process based on the dynamic time warping algorithm, the following symbols are defined for further explanation. Assuming the reference template is $R = \{r_1, r_2, r_3, r_4, r_5, r_6,\}$, r_m represents the mean of 6 attributes and r_6 represents their standard deviation. Likewise, the test template is $T = \{t_1, t_2, t_3, t_4\}$, with t_m representing the mean of 4 attributes and t_3 representing their standard deviation.

The specific process of template matching is as follows:

$$\begin{aligned} r_i^* &= \frac{r_t - r_m}{r_s} i \in,\\ t_j^* &= \frac{t_j - t_m}{t_s} j \in . \end{aligned} \tag{12}$$

5. Results

Included in the CAIDA dataset is anonymized passively monitored traffic from the University of Chicago's Equinix high-speed Internet backbone. The experimental results are shown in Figure 4.

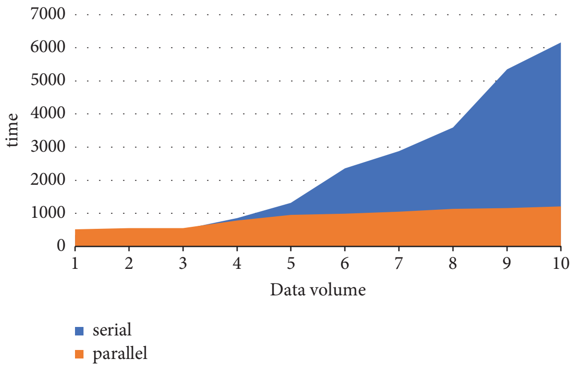

Figure 4: Comparison of parallel and serial time.

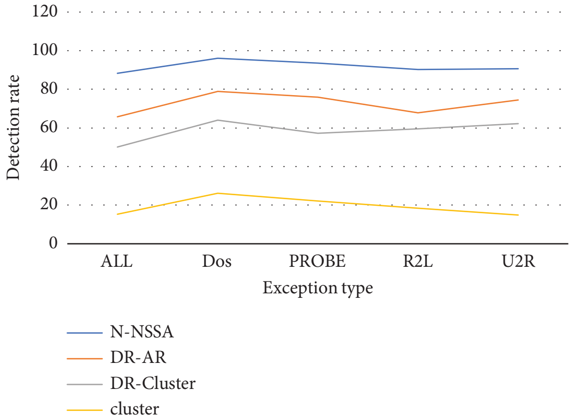

Figure 5: Comparison of detection rates of four methods.

The same experimental environment was used with the same data set and the same four algorithms included in the model. The results of the experiments are shown in Figure 5. In the first method, only the distributed parallel clustering algorithm was used to analyze the experimental data; in the second method, the data were clustered after adding a dimensionality reduction operation to the data. In this way, the detection rate increased significantly. Dimensionality reduction of the data is very effective. Therefore, the third method uses a dataset with the same dimensionality reduction and analyses the dataset by association rules. The fourth method is a security situation recognition model by using neural networks. First, the data are cleaned and analyzed for situational awareness. The neural system is used to correct the results of the preliminary analysis and to obtain the final results.

In order to verify the acceleration of distributed parallelization model, 1–5 nodes were selected as the experimental cluster in this experiment. The largest data in the "parallel and serial time comparison" experiment were selected as the test data set.

Figure 6 shows the number of nodes continues to increase, and the speedup curve of the model also increases. However, when the number of nodes is 3 to 4, the growth rate of the curve becomes slower. This is because the increasing number of nodes increases the cost of communicating with each other. Communication between nodes will consume certain resources and time. It can be seen from experiments that the security situational awareness model based on neural performs well in terms of accuracy, false detection rate, time, and efficiency.

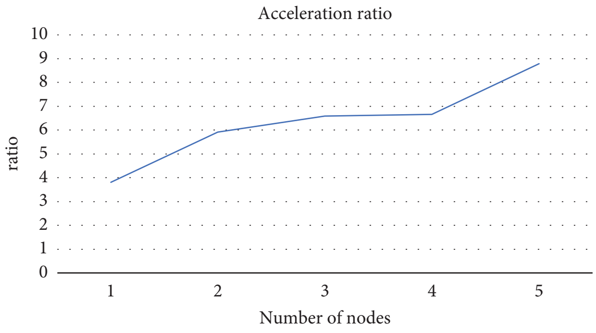

Figure 6: Acceleration ratio of different nodes.

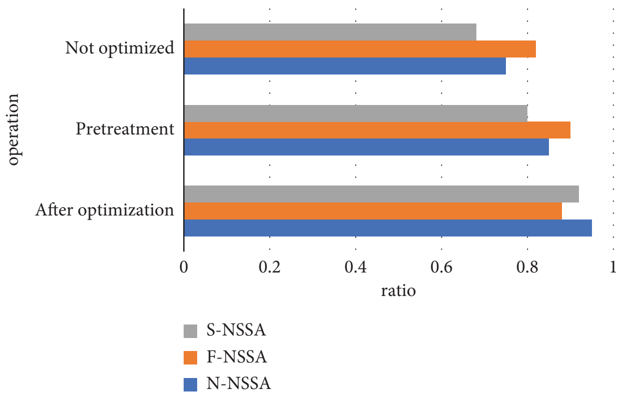

Figure 7: Comparison of detection rates of three models.

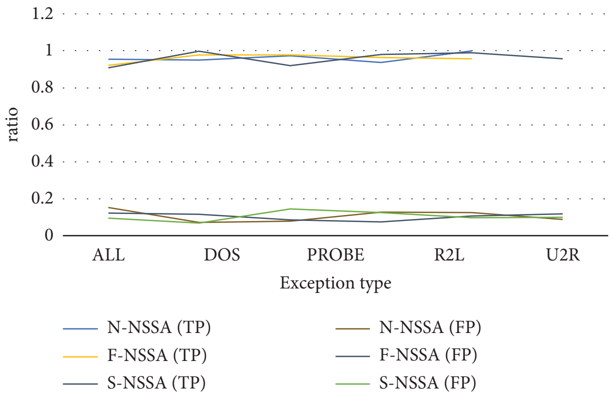

Figure 8: Comparison of detection rate and false detection rate of three models.

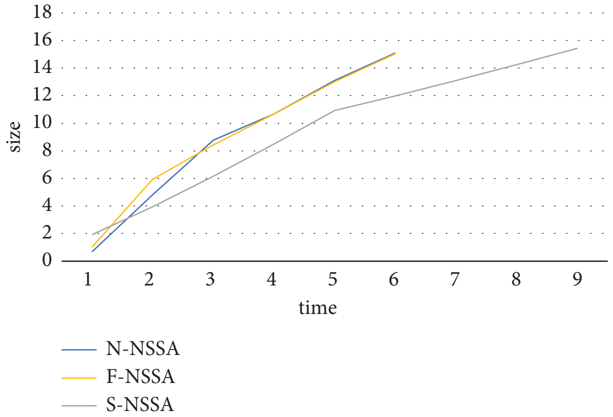

Figure 9: Comparison of different data scales.

In Figure 7, three models proposed in security situational awareness have been greatly improved compared with experiments in the first part, and accuracy rates are all above 90%. However, different from the first two parts, in this part of the experiment, the detection rate of neural and star structure surpassed the detection rate of the forest model. This is because original data are effectively preprocessed after the first stage of situational awareness, so the neural model and star structure model overcome sensitivity to shortcomings of the data set, so the detection rate is higher.

In Figure 8, the second indicator is false detection rate (That is, abnormal data are incorrectly detected. The proportion of data judged to be normal.). From the data in the figure, it is not difficult to find that the detection rate of the three models is above 90%, and each abnormal type can be detected, and the false detection rate is below 10%. Therefore, the three models proposed are used for security situational awareness performance is better.

From Figure 9, we can obtain the conclusion that when dealing with datasets of the same size. This conclusion is justified in the third part of the first set of experiments. Due to model characteristics of the star structure, although peripheral modules are processed in parallel by many nodes at the same time, all data in the situation projection stage are processed by the central core module, which leads to performance in terms of time efficiency. Not as good as the forest model. Each node performs the tree-building operation of part of the data, thereby forming a forest model of the entire security situation. Finally, the final security situation is judged by the mode of the result of each tree, which not only avoids one-sidedness but also makes the processing of data very efficient due to such structural characteristics.

As the number of nodes increases gradually, the processing time between models is compared. In Figure 10, the scale of the cluster continues to expand, and the number of nodes participating in parallel operations gradually increases, but time consumption for data communication and transmission between nodes will also increase. The node acceleration ratio of the neural model is relatively low because communication between each node is relatively large during the error adjustment learning process of the neural network. When the scale of the cluster expands, the number of nodes processing data in parallel also increases, which makes improvement of processing efficiency not obvious when the number of nodes increases. In this part of the experiments, node speedup ratios of the forest model and star structure are comparable. As mentioned in the previous part, due to the structural characteristics of the forest model itself, in the whole process of security situation awareness, each node is used for data processing, such as the construction of a decision tree independently and data transmission between each other is relatively small, so the acceleration ratio is very good. The star structure is the same

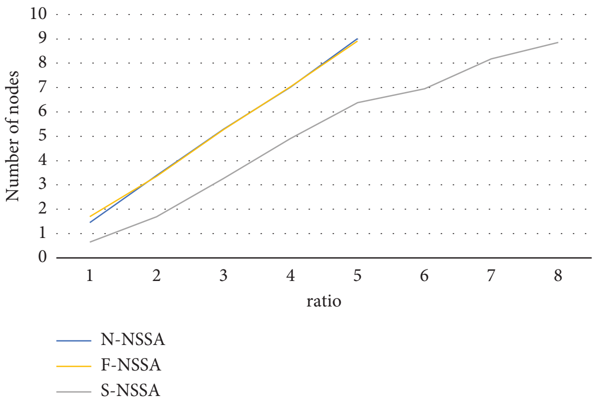

Figure 10: Acceleration ratio of different nodes.

as the forest model in the first two stages of security situational awareness. Each node processes data independently, and the parallel effect is good. Only in the third stage will there be data transmission between nodes.

Figure 11 shows the rising curve is the acceleration ratio curve, and the falling curve is the scale ratio curve. Both curves are approximately linear. The experimental results show that the model can achieve a good speedup ratio and scale ratio in a distributed cluster.

6. Countermeasures and Proposals to Strengthen Network Security Governance in the Era of Big Data

6.1. Improving Legislation, Systems, and Legal Awareness of Network Security. The law is a system of mandatory specifications; it is an institutional guarantee for network security management. In the absence of law, even the most advanced technologies and management tools are difficult to manage really well. Laws and regulations related to network security clearly warn all people in different ways about what behavior is unacceptable and what is allowed online. Although the law is a prevention tool and an enforcement tool, the reason for network security is to create the strongest and most reliable line of defense.

6.1.1. Effective Laws and Regulations Related to Network Security Improve Data Management. From the first "Regulations for the Protection of Computer Information System Security" adopted and enacted in 1994 to the first major network security law, the "Network Security Law of the People's Republic of China" adopted and enacted in 2017, the government has paid sufficient attention and firm support to network security management in terms of policies and regulations.

However, the development of the Internet is inexorable, and any technological breakthrough may lead to radical changes. Legislation, only steady and over time to master the dynamics and direction of the development of the Internet, insisting on the work of Internet security management at the level of legal support, timely and effective for the emerging things, to standardize the management system and gradually diversify the data—the scope of the network industry lifeline management system, so that only in this way can we carefully ensure the healthy, orderly, and safe development of Kit's network industry.

6.1.2. Raising Public Awareness and Promoting Network Security Laws and Regulations. Although the network provides people with many conveniences, it also brings some network security issues. In order to ensure the network security of every citizen, it is necessary to raise legal awareness, network security awareness, master basic protection skills to prevent Internet fraud and theft, improve the quality of civilized Internet access, and consciously regulate lawful Internet behavior. While each can work well on their own, they can come together in a strong united force for network security.

Since 2014, public safety agencies across the country have held annual on-site promotional events in September, "National Network Security Awareness Week—Rule of Law Day," to effectively raise everyone's awareness of safe Internet use. The Internet Security Department of Taiyuan Public Security Bureau also organized large-scale promotional activities during the annual Network Security Publicity Week and gradually developed from the May Day market in different areas of the city, various mass convergence of legal advertising, antifraud advertising, answering questions, and solving problems.

6.1.3. Establishing a Data Resource Security Protection Model Suitable for Taiyuan. At present, China's big data industry is mainly concentrated in developed regions such as North, Guangzhou, and Shenzhen, and many famous enterprises such as Sina, Baidu, 360 and Tencent have brought top scientific and technological talents to these regions, making these regions leaders in the big data industry. In addition, the southwestern circle of the big data industry, centered in Guizhou, has been studied and established as a regulatory policy system conducive to promoting big data innovation and development and actively introduced big data-related key enterprises and talents, which has led to significant development and gradually created a "Chinese big data center."

In the traditional region, Shanxi is disadvantaged due to regional constraints, transportation, climate, and other factors; since Shanxi is also in an important period of transformation and development, the demonstration zone of comprehensive reform is also located in Taiyuan. I believe that Taiyuan should increase economic investment in big data, actively introduce multisector talents, seek to seize development opportunities, continue to carry out research and development, achieve technical and economic development, and realize the development of the industry.

6.2. Strict Regulation of Internet-Related Crimes and Cleaning Up the Network Environment. General Secretary Xi Jinping attaches great importance to work on network security, stressing that "cyberspace is the common intellectual home

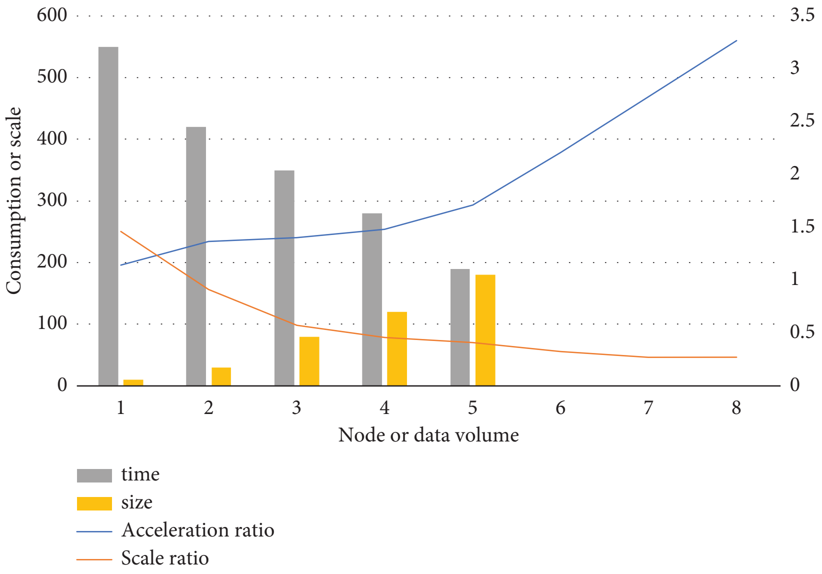

Figure 11: Acceleration ratio and scale ratio.

of hundreds of millions of people," which places specific requirements on public security organs at all levels to strictly manage network security and crack down on illegal and criminal activities using it. Public security institutions, especially network security departments, should keep in mind their purpose to serve citizens wholeheartedly and depart from their responsibilities and actions to implement a mechanism for the rapid resolution of Internet-related cases and incidents to create a safe and transparent network environment for citizens.

6.2.1. Speeding Up the Process and Efficiency of Security Management Work and Timely Follow-Up and Supplement Relevant Measures. Due to the regional mandate of public security authorities, many Internet cases are handled in different regions, requiring cooperation between public security authorities in different locations. In order to speed up the process of network security management and improve the effectiveness of network security management, it is necessary to strengthen interconnection mechanisms among regional public security authorities, especially in the field of network security. Regional public security agencies with an advantage in the "big data" field should share data resources to help fight and punish network attacks, while agencies in regions that are relatively behind should strengthen the capabilities of their personnel in the investigation, prevention and control, detection and surveillance, etc. Big Data" will only be available when talent and data are combined. Only a combination of talent and data can serve as "big data" in this domain.

National network security should have the idea of "one chess piece" and resolutely break the regional barriers that can set the Ministry of Public Security, the Provincial Public Security Bureau as the main battlefield for grabbing and the Municipal Public Security Bureau as the main battlefield for the network security situation, and perform more "cleaning the network," "fighting pornography and illegal," etc. More special operations for network security, such as "cleaning up the network," "fighting pornography and illegal activities," etc., should be implemented, training teams based on real-world operations, drawing vitality from experience and lessons learned, and leveraging the advantages of network security information resources to fight cybercrime and protect network security as protection as a powerful force in the fight against cybercrime and in the protection of network security.

6.2.2. Maximizing the Use of Big Data Resources to Develop and Implementing Proactive Network Security Management. "In order to fully leverage big data, we must first look at the weaknesses of network security management itself, starting with ourselves.

Given the difficulty of managing network security, managing network access rights, strengthening data encryption, and hardening terminals while not being lost as effective means of protecting data security, these are things that cannot be carried out once and for all or given the evolution of "big data." Therefore, by fully utilizing the resource of "big data," implementing and executing future-oriented network security management, improving the technical level of network security administrators, checking and filling time gaps, and updating firmware, etc., the network security management must be smoothly implemented to overcome attacks on the network.

6.2.3. Improving the Early Warning Function of Big Data and the Importance of Decision-Making. So-called "big data" is not big data but data analysis based on big data with strong predictive capabilities. The already mentioned "high school girl pregnancy prediction" is just one small application of "big data.

When Internet-related incidents and accidents occur frequently, it is necessary to use big data's early warning capabilities to prepare a response, not only to monitor and detect incidents and accidents after they occur but also to alert criminals, identify possible criminal activities in advance, and eliminate crime in its cradle.

6.3. Strengthening Software and Hardware in the Network Security Management Sector.

"Without good 'weapons' in the age of big data, it is certainly impossible to win the 'battle' of network security management. I believe that the software and hardware of network security management departments should be strengthened on the following three fronts.

6.3.1. Use of High-Precision Hardware and Increased Investment in Technology.

With "big data" in the total number of ZBs and an average annual growth rate of more than 25%, conventional electronic equipment is unlikely to perform the important tasks of security management and data analyses, so there is a need to invest vigorously in research and development. Big data is a strategic resource that concerns the rights and interests of individual citizens and national security and must be fully protected. Technologies and equipment in the research and development phase must be fully tested in both real-world and laboratory simulation trials.

6.3.2. Data Discovery, Removal of Data Barriers, and Data Sharing.

Data are considered a resource and a wealth, but they should never be private property. If existing data regions and industry barriers are reasonable in terms of data security and protecting commercial interests, the future trend should be toward data sharing and collaborative development.

I think the priority should be to remove barriers in two directions. The first is the direction of national security, and people's lives should be coordinated, resources should be allocated rationally, all data should be analyzed together, and the Chinese system should be better in all aspects so that it contributes more to the direction of people. Second, the direction of combat and punishment should also be the aggregation of data, development and sharing, in the spirit of the "one defeat" idea, to achieve the maintenance of national security.

6.3.3. Intensifying Efforts to Train Big Data Professionals.

In China, network security, big data pools, and other areas of talent are relatively scarce. In the face of the urgent demands of the big data era, we must put more effort into cultivating big data professionals. Talent is an important embodiment of national strength, and the lack of a talent pool is a weakness. The global state of big data is currently losing momentum, and we need to pick up the pace to avoid being left behind.

I think we can improve our big data talent pool in three ways. First, the state should add professional training programs in network security, big data, and other related fields to human resources education, especially at the undergraduate and graduate levels, based on sustainable, long-term training of professionals. Second, regardless of the model, we combine the reality of talent recruitment and improve the ability of the elite talent in society to develop, analyze, and use big data. Third, from the real work, we increase funding for talent training and encourage innovation. Third, from the real work, we increase investment for talent training and cultivate practical talent, with three aspects of joint efforts to strengthen China's network security management, big data, and other areas of common strength.

7. Conclusion

Information security education is an important part of quality education in China. Improving public information security education is very important for the development of society itself and the maintenance of national security. In recent years, the government has introduced a series of measures and policies to strengthen the research and development of information security technology and promote the development of information security protection levels; universities gradually pay attention to information security education and carry out related academic work, and the public awareness of information security education has been increasing, and information security education in China has been supported by a good environment and protection. Based on the current situation of information security education in China, this paper clarifies the concept and characteristics of information security education in the era of big data and then proposes an adaptive security situational awareness model to cope with the problems brought by streaming data in a complex environment based on the general theory of previous research. The model uses a self-learning strategy with error feedback for adaptive learning, which effectively reduces the probability of error detection and increases the detection rate to more than 98%, enabling it to be extended to large-scale processing of high-dimensional streaming data.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The study was supported by the Ministry of Education of Science and Technology Development Center of Production Innovation Fund China's Colleges and Universities (Grant no. 2020 IT A07027)".

References

[1] M. Du, "Application of information communication network security management and control based on big data technology," *International Journal of Communication Systems*, vol. 35, no. 5, Article ID e4643, 2022.

[2] A. Alharthi, V. Krotov, and M. Bowman, “Addressing barriers to big data,” *Business Horizons*, vol. 60, no. 3, pp. 285–292, 2017.

[3] J. Gao, “Analysis of enterprise financial accounting information management from the perspective of big data,” *International Journal of Science and Research*, vol. 11, no. 5, pp. 1272–1276, 2022.

[4] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang, “Big data service architecture a survey,” *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.

[5] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, “Big data analysis-based secure cluster management for optimized control plane in software-defined networks,” *IEEE Transactions on Network and Service Management*, vol. 15, no. 1, pp. 27–38, 2018.

[6] S. Tiwari, H. M. Wee, and Y. Daryanto, “Big data analytics in supply chain management between 2010 and 2016: insights to industries,” *Computers & Industrial Engineering*, vol. 115, pp. 319–330, 2018.

[7] Q. Feng and J. G. Shanthikumar, “How research in production and operations management may evolve in the era of big data,” *Production and Operations Management*, vol. 27, no. 9, pp. 1670–1684, 2018.

[8] Y. Zhang, J. Ren, J. Liu, C. Xu, H. Guo, and Y. Liu, “A survey on emerging computing paradigms for big data,” *Chinese Journal of Electronics*, vol. 26, no. 1, pp. 1–12, 2017.

[9] Y. Zhang, T. Huang, and E. F. Bompard, “Big data analytics in smart grids: a review,” *Energy informatics*, vol. 1, no. 1, pp. 8–24, 2018.

[10] K. Wang, Y. Wang, X. Hu et al., “Wireless big data computing in smart grid,” *IEEE Wireless Communications*, vol. 24, no. 2, pp. 58–64, 2017.

[11] W. Xu, H. Zhou, N. Cheng et al., “Internet of vehicles in big data era,” *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 19–35, 2018.

[12] F. Ullah and M. A. Babar, “On the scalability of big data cyber security analytics systems,” *Journal of Network and Computer Applications*, vol. 198, Article ID 103294, 2022.

[13] H. Xiao, *Information Security Management of Smart Campus System Based on Big Data*, Forest Chemicals Review, 2022.

[14] A. G. Sreedevi, T. Nitya Harshitha, V. Sugumaran, and P. Shankar, “Application of cognitive computing in healthcare, cybersecurity, big data and IoT: a literature review,” *Information Processing and Management*, vol. 59, no. 2, Article ID 102888, 2022.

[15] W. Dai, L. Qiu, A. Wu, and M. Qiu, “Cloud infrastructure resource allocation for big data applications,” *IEEE Transactions on Big Data*, vol. 4, no. 3, pp. 313–324, 2018.

[16] Z. Chang, L. Lei, Z. Zhou, S. Mao, and T. Ristaniemi, “Learn to cache machine learning for network edge caching in the big data era,” *IEEE Wireless Communications*, vol. 25, no. 3, pp. 28–35, 2018.

[17] D. Arunachalam, N. Kumar, and J. P. Kawalek, “Understanding big data analytics capabilities in supply chain management: unravelling the issues, challenges and implications for practice,” *Transportation Research Part E Logistics and Transportation Review*, vol. 114, pp. 416–436, 2018.

[18] R. A. Ariyaluran Habeeb, F. Nasaruddin, A. Gani, I. A. Targio Hashem, E. Ahmed, and M. Imran, “Real-time big data processing for anomaly detection a survey,” *International Journal of Information Management*, vol. 45, pp. 289–307, 2019.

[19] G. S. Sriram and G. S. Sriram, “Security challenges of big data computing,” *International Research Journal of Modernization in Engineering Technology and Science*, vol. 4, no. 1, pp. 1164–1171, 2022.

[20] Q. Qi and F. Tao, “Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison,” *IEEE Access*, vol. 6, pp. 3585–3593, 2018.

[21] Z. Cai, Y. Liu, B. Tang et al., “Dynamics of minimal residual disease defines a novel risk-classification and the role of allo-HSCT in adult Ph-negative B-cell acute lymphoblastic leukemia,” *Leukemia and Lymphoma*, vol. 34, no. 3, pp. 1–10, 2022.

[22] W. Qi, M. Sun, and S. AghaSeyedHosseini, “Facilitating big-data management in modern business and organizations using cloud computing: a comprehensive study—Corrigendum,” *Journal of Management and Organization*, vol. 127 pages, 2022.

[23] J. Li, “Venture financing risk assessment and risk control algorithm for small and medium-sized enterprises in the era of big data,” *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 611–622, 2022.

[24] S. Sai Kumar, A. R. Reddy, B. S. Krishna, J. Nageswara Rao, and A. Kiran, “Privacy preserving with modified grey wolf optimization over big data using optimal K anonymization approach,” *Journal of Interconnection Networks*, vol. 2022, Article ID 2141039, 2022.

[25] H. Xiao, *Exploration of Network Information Security Technology and Prevention in the Digital Age*, Forest Chemicals Review, 2022.

[26] T. A. T. Ali, “Geospatial big data analytics applications trends, challenges opportunities,” *Asian Basic and Applied Research Journal*, vol. 15 pages, 2022.