Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 3141568, 13 pages
https://doi.org/10.1155/2022/3141568

Research Article

Security Attitude Prediction Model of Secret-Related Computer Information System Based on Distributed Parallel Computing Programming

Ling Sun[1] **and Dali Gao**[2,3]

[1]*School of Information Engineering, Henan University of Animal Husbandry and Economy, Zhengzhou 450044, Henan, China*
[2]*School of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou 362000, Fujian, China*
[3]*Key Laboratory of Intelligent Computing and Information Processing, Fujian Province University, Quanzhou 362000, Fujian, China*

Correspondence should be addressed to Dali Gao; laogaogao@qztc.edu.cn

Received 18 January 2022; Revised 28 February 2022; Accepted 4 March 2022; Published 25 March 2022

Academic Editor: Gengxin Sun

In recent years, there has been an upward trend in the number of leaked secrets. Among them, secret-related computers or networks are connected to the Internet in violation of regulations, cross-use of mobile storage media, and poor security management of secret-related intranets, which are the main reasons for leaks. Therefore, it is of great significance to study the physical isolation and protection technology of classified information systems. Physical isolation is an important part of the protection of classified information systems, which cuts off the possibility of unauthorized outflow of information from the network environment. To achieve the physical isolation of the network environment and build a safe and reliable network, it is necessary to continuously improve the level of network construction and strengthen network management capabilities. At present, the realization of physical isolation technology mainly relies on security products such as firewall, intrusion detection, illegal outreach, host monitoring, and auditing. This study analyzes network security systems such as intrusion detection, network scanning, and firewall. Establishing a model based on network security vulnerabilities and making up for network security hidden dangers caused by holes are generally a passive security system. In a network, the leader of network behavior—human behavior—needs to be constrained according to the requirements of the security management system and monitoring. Accordingly, this study proposes a security monitoring system for computer information network involving classified computer. The system can analyze, monitor, manage, and process the network behavior of the terminal computer host in the local area network, to achieve the purpose of reducing security risks in the network system. Based on the evaluation value sequence, the initial prediction value sequence is obtained by sliding adaptive triple exponential smoothing method. The time-varying weighted Markov chain is used for error prediction, the initial prediction value is corrected, and the accuracy of security situation prediction is improved. According to the security protection requirements of secret-related information systems, a complete, safe, reliable, and controllable security protection system for secret-related information systems is constructed, and the existing security risks and loopholes in secret-related information systems are eliminated to the greatest extent possible. This enables the confidentiality, integrity, and availability of confidential data and information in the computer information system to be reliably protected.

1. Introduction

The rapid development of network technology has brought great convenience to human production and life, and at the same time, the scale of the network is expanding, the network environment is becoming more and more complex, and the network security issues have brought great challenges to human beings. Under the continuous promotion of information and intelligence of social life, information system as a carrier of information in the network, its scale and environment are increasingly large and complex, and because it contains huge commercial

interests and other values, it comes with more frequent and variable attack techniques [1]. Since information systems are the infrastructure for the continuous promotion of informatization and intelligence in all sectors of society, the normal work operation of organizations ranging from large countries to small commercial organizations depends on information systems, and once their security is compromised, it will bring irreparable and huge losses. The security posture assessment model is necessary for information system security managers to obtain the dynamic security situation of the system, determine abnormal system events, and make reasonable decisions, while the security posture prediction is based on the results of the security posture assessment, which has a positive effect on preventing high-risk events and improving the emergency response capability of information systems. Due to the supporting role of information systems in social life, how to enhance the security defense capability of information systems and improve the level of information security risk control has become a major issue that needs to be addressed. In the face of the increasing number of network attacks, academia has proposed many defense measures for information system security, but these defense measures mainly use the patching method, focusing on after-the-fact defense, but in the real network environment when the risk is found before the network has already been attacked, it results in information leakage, tampering, and other hazards. How to defend against possible cyberattacks in advance and reduce the probability of cyberattacks on network information systems has become a focus of attention.

The diversity of parallel platforms lies in the number and types of parallel platforms. With the rapid development of the trend of parallelization of computer processors and architectures in recent years, a variety of multicore processor calculators, Kor-terminated servers, and gas pedals of various architectures and systems constitute various parallel computing platforms, and with the increase in the number and diversity of various parallel computing platforms and parallel programming models, it is necessary to have the ability to quickly generate parallel programs for the required target parallel platform [2]. In this study, by studying the parallelization of community discovery algorithms using distributed computing platforms, the adaptive weighted threshold scheduling algorithm is calculated and implemented to form the information system security protection initiative, and the security posture assessment of information systems is performed by acquiring security elements in a certain time and space and integrating and analyzing the acquired data information to determine the current system security status [3]. The security posture assessment model is necessary for information system security managers to obtain the dynamic security situation of the system, determine abnormal system events, and make reasonable decisions, while the security posture prediction is based on the results obtained from the security posture assessment, which has a positive effect on preventing high-risk events and improving the emergency response capability of information systems.

2. Related Work

In cyber security incidents, the basic characteristics of information system security risks are related to simultaneous offensive and defensive actions. Therefore, it is of good research significance to study security risk assessment methods based on comprehensive consideration of the role of attack and defense. Ibrahim [4] uses letter entropy to quantify and calculate the gain situation of offensive and defensive behaviors and establishes a game analysis model to study the information offensive and defensive confrontation process to study the security risk situation of privacy attacks. Tykkyläinen and Ritala [5] demonstrated the reliability and validity of information risk assessment through stochastic Petri nets in game theory. The literature [6] analyzes the security of information systems in the context of wireless networks by constructing a game model to improve the routing protocols based on the analysis of their information throughput using the Nash equilibrium. Goni et al. [7] proposes a distributed framework based on Hadoop and improves the Apriori algorithm using the Boolean matrices to replace the original transactional database and then uses many operations on the matrix and other logical operators to find frequent patterns and finally divides the matrix into multiple parts for parallel computation. Literature [8] proposed an association rule mining algorithm based on MapReduce to convert each row of input transactions into the binomial format. Moroz and Gamble [9] tries to use hash tables and hash trees for the candidate set storage structure in MapReduce-based implementation of Apriori algorithm, and experiments find that the speed of association rule mining is further improved. The literature [10] proposed a maximal AprioriMR algorithm for compressing frequent patterns and an algorithm for censoring the search space using known anti-monotonicity, both of which have higher computational efficiency for frequent set mining. The literature [11] proposes a new distributed framework called Spark using its in-memory computing technique with the help of resilient distributed dataset (RDD) storage. The I/O communication on RDD is very efficient and fast, and RDD stores the results in the main memory at the end of the iteration and makes them available for the next RDD that stores the result in the main memory at the end of the iteration and makes it available for the next iteration. The literature [12] proposed a distributed frequent itemset mining algorithm DFIMA for big data analysis using matrix-based pruning technique to reduce the candidate size, and experiments show that Spark-based frequent set mining algorithm has good computational efficiency.

To improve the accuracy of the Bayesian model for system security assessment, the literature [13] proposed a security risk assessment model based on the Bayesian network, which is based on the Bayesian theory and combines conditional probability with ordered weighted average operator to analyze system threats and thus achieve quantitative assessment of system security risk. The literature [14] proposed a system intrusion detection method based on a parameterless Bayesian model to monitor network activities and thus protect the system from attacks and identify the types of network attacks

using classification, which introduces a Gaussian mixture method to solve the overfitting problem and has high detection accuracy. The literature [15] proposes a dynamic risk assessment model for information security of industrial control systems based on the Bayesian attack graphs, which combines the prior distribution and real-time attack sample data obtained from intrusion detection systems and uses the Bayesian parameter learning to dynamically adjust the conditional probabilities of nodes to achieve dynamic risk assessment of the overall security of the target network. To apply the Markov models to security risk analysis, Omer et al. [16] propose a Markov model-based security assessment method for computer systems to analyze the threat probability to the system by calculating the system state probability and to study the dependence of the system security state probability on the threat probability. Finally, the relaxation time is introduced and the range of system protection parameters is determined on this basis. Randhawa et al. [17] proposes a system operation detection model based on the Markov decision process, using incremental construction methods to optimize the performance of the system during operation and to analyze the security risks faced by the system. The literature [18] proposes a method for calculating the probability distribution of security threats based on Markov chains and a generic vulnerability scoring system and uses the security threat model to analyze the likelihood of the system suffering from attack type. Malallah H S proposed a Markov model-based system security assessment method to improve the acquisition of observation sequences and use them to represent network security and assess risk using state transfer matrix to improve the rationality and accuracy of risk assessment.

3. Research on the Key Technology of Distributed Parallel Computing Programming

3.1. Distributed Parallel Computing Efficiency Study. Parallel computing is opposed to serial, where the main feature of the serial computing model is that a problem is decomposed into some discrete instructions that are executed sequentially on a single processor, and at most one instruction is executed at any given time. Parallel computing decomposes a problem to be solved into multiple subtasks, which are assigned to different processors, and each processor cooperates to execute the subtasks in parallel, thus achieving the purpose of speeding up the solution or increasing the size of the solution. Concurrent computing can use multiple computing resources simultaneously to solve a computational problem. Computational resources in this context refer to computers with multiple processors or multicore computers, clusters, or fleets. Parallel computing requires three basic conditions to be carried out: first, the platform is parallel. That is, it contains at least two processors, multiple computers, or multiple nodes that are interconnected and communicate with each other. Second, the problem to be solved can be parallelized. The problem to be solved can be decomposed, and the final results can be combined. The last is parallel programming, that is to design parallel algorithms based on specific tasks on a parallel platform, and then writing parallel programs to realize parallel computing [19]. Domain decomposition is the decomposition of data according to the problem, with each parallel task manipulating a part of the data; functional decomposition focuses on the computation to be performed rather than on the data manipulated by the computation, and the problem is decomposed according to the content of the computation, with each task performing a part of the entire job, ideal for multitasking problems. The difference between parallel computing and distributed computing is that distributed computing organizes distributed computing units to collaborate to coordinate resource access and improve resource utilization, while parallel computing emphasizes increasing the speed of solution scale of solving the same problem.

A cluster is a group of independent computers interconnected by a high-speed network that forms a group and is managed in a single system model. When a client interacts with a cluster, the cluster can be seen as an independent server. Clusters are the most popular high-performance computing platform because of their scalability, ease of parallel programming, high computational performance, and high reliability. The cluster has a dedicated job scheduling system to efficiently manage the various resources in the system and the jobs submitted by the users. The purpose is to make full use of the hardware and software resources and valuable CPU time of the cluster so that the system has a high-throughput rate and utilization. Users are required to submit scripts in a specific format for the system to allocate resources and execute computing tasks, as shown in Figure 1.

To conveniently describe the performance and effectiveness of parallel computing, the acceleration ratio metric is generally used for metrics. The acceleration ratio is calculated as follows:

$$S_{\mathrm{k}} = \frac{P_x - P_\mu}{\sigma}, \tag{2}$$

where P_μ is the serial execution time and P_x is the parallel execution time using x one processor, and the efficiency of the parallel algorithm is further obtained from the acceleration ratio.

$$\eta = \frac{S_k}{x} + c. \tag{3}$$

For a given computational problem, assuming that the percentage of the serial is m, the parallel speedup ratio using x one processor is as follows:

$$S_k = m \cdot \sum_{i=1}^{n} \frac{\left(P_i - P_\mu\right)}{x}. \tag{4}$$

As m increases, S_k also increases, but it is upper bounded. That is, the speedup multiplier cannot exceed $1/x$, regardless of the number of processors used—the speedup potential of a program is determined by the proportion of its parts that can be parallelized, and it is worth noting that this formula is idealized and does not take into account process overhead and locks.

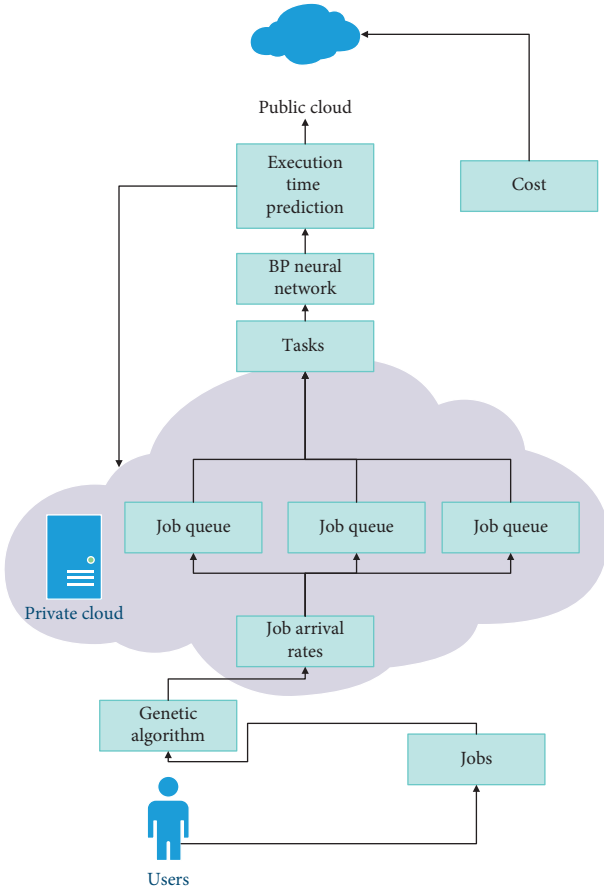

Figure 1: Cluster job scheduling model.

GNSS parallel data processing refers to the GNSS data processing theory and algorithm process based on GNSS data processing; combined with parallel computing technology, it will be parallelized processing design and configured to parallel computer processing. In GNSS parallel data processing, to give full play to the effectiveness of parallel computing, the specific parallel method needs to combine the data processing mode, scale, and process to parallelize and optimize the computational tasks [20]. First, the computation task is decomposed according to the number of cores or nodes involved in the computation and the size of the computation task; then, the subtasks are assigned to each process for parallel computation; finally, the computation results are returned to the main process for combination to obtain the final results. There are usually two basic methods used for parallel task decomposition: domain decomposition and functional decomposition. Domain decomposition decomposes the data according to the problem, and each parallel task manipulates a part of the data; functional decomposition focuses on the computation to be performed rather than the data manipulated by the computation and decomposes the problem according to the content of the computation, and each task performs a part of the whole work, which is ideal for multitask problems. In GNSS parallel computing design, both domain decomposition and functional decomposition are involved; e.g., the path of data decomposition is mainly utilized when simply considering large-scale matrix operations, while task decomposition methods are used when performing the system of normal equations construction or data preprocessing part, or both. GNSS parallel data processing needs to be based on parallel programming models, and the common parallel programming models include shared memory model, thread model, message-passing model, hybrid model, and multi-program multi-data model. In this study, the parallel programming model mainly uses shared memory, multi-threading, message passing, and hybrid models.

3.2. Parallel Computing Task Scheduling Model. The description of parallel computing tasks (including but not limited to MapReduce) in the Hadoop ecosystem, including the description of the task itself, the description of the cluster, and the description of the state of the user-submitted jobs, specifically describes some key parameters that affect performance. When the standard deviation of the actual value series of security posture is small, the fluctuation of security posture is small, and the predicted impact of recent and distant data on the value of security posture is approximately the same, so when the sliding window width is larger, the time span is large, and the number of historical data contained in the window is increased, which facilitates the calculation to obtain a more reasonable static smoothing coefficient and reduces the error of the initial predicted and

actual values. A Hadoop cluster will contain a set of physical hosts of different types, the main difference lies in hardware resources, and we define each physical host as PM, which is of type Type, randomly distributed on the Rack Rack, and the racks are connected with a high-performance network, and then, the following cluster definition can be obtained according to the actual relationship between the cluster and the physical machines.

$$\text{Cluster} = \{PC_1, PC_2, PC_3, \ldots, PC_i, \ldots, PC_M\}. \quad (5)$$

As shown in equation (4), there are K physical hosts PM in the cluster, each physical host has multiple CPU cores, a certain size of memory RAM, hard disk space, and a network band link for uploading and downloading data, the same physical machine will have homogeneous CPU cores and memory, and different physical machines may have differences in core performance or memory performance, so the whole cluster is heterogeneous The physical hosts of the whole cluster are randomly distributed on different rack racks. When a user designs an application to process the data in the cluster, we call it a Hadoop job submitted, and we define a job in Hadoop as follows:

$$\text{Job} = \{TK_1, TK_2, TK_3, \ldots, TK_i, \ldots, TK_M\}. \quad (6)$$

In the above equation, K represents the specific task to be processed by each job, and T indicates that the job reaches the cluster, i.e., the point in time when the application starts running, i.e., when the user submits the job. Each job should also contain attributes such as the amount of data and computation required for the job to process.

The Hadoop ecosystem is a computing cluster with heterogeneous computing resources, and these resources are shared by each user logged into the system without applications that take up computing resources for long periods. The goal of the full paper is to design an efficient task scheduling strategy to fit across the heterogeneous cluster so that the task completion time is as small as possible in the heterogeneous case. Generally, distributed parallel computing task scheduling is affected by many factors, such as task start time, because there is a dependency between Map tasks and Reduce tasks in Hadoop; i.e., in general, the input data required by Reduce tasks are not fully prepared until the execution of Map tasks is completed. The output of different Map tasks will correspond to the input of specific. In this study, we assume that the corresponding Reduce program will be started only after all Map tasks are processed to simplify the overall operating model. Further, the following assumptions are also made for this purpose, and we will describe them and analyze the possible implications.

Assuming that the current Hadoop cluster consists of 15 physical machines in 3 racks, Rack1, Rack2, and Rack3, each rack has an average of 5 physical machines as compute nodes as shown in Figure 2, ignoring the similarities and differences between NameNode and DataNode as described above. Let us look at it this way: when a user submits a job with only one task and the execution time of that task is variable, when the Task is assigned to the node whose data it needs (assuming that Node3 has both the Task program code segment and the data it needs for processing), the execution time of the Data-Local type is 10 units of time, and the total task execution time is 10 units of time. When the task is assigned to another node in the same rack as the node where the data it needs are located, the total time can ignore the data transfer time if the data transfer time is short and the processing time is long (e.g., Node8, Rack-Local1 type). When the task is assigned to a rack that is not in the same rack or even in the same data center as the data copy (Node13, Remote1 type vs. Node14, Remote2 type), the completion time increases significantly and the data transfer time can be at worst more than 90% of the total completion time, which is a situation that must be avoided. Whether it is traditional data mining or data analysis in the big data environment, clustering as a basic process to automatically categorize unknown basic process of automatic categorization of data, it can be used in both the preprocessing stage of data and data mining processing and is different from classification. The study of data locality and resource allocation is of great importance and value to improve the overall cluster performance.

When a distributed computing cluster is started, the cluster should consist of M computing nodes, i.e., physical machines, and we consider that the computing resources available for each physical machine are not certain, but the resources available in the cluster at any given time are available according to the monitoring procedure, and then, we define the computing node resource availability matrix of a distributed cluster containing M physical computing nodes as follows [21]:

$$\text{Resource} = \begin{bmatrix} \text{CPU}_1 & \ldots & \text{CPU}_i & \ldots & \text{CPU}_M \\ \text{RAM}_1 & \ldots & \text{RAM}_i & \ldots & \text{RAM}_M \\ \text{Disk}_1 & \ldots & \text{Disk}_i & \ldots & \text{Disk}_M \end{bmatrix}. \quad (7)$$

Here, we only consider the CPU, memory RAM, and hard disk capacity of each physical machine, corresponding to the computational task resource demand matrix, and each task demanded resource is provided by the physical node with the corresponding dimension of available resources, which provides convenience for subsequent algorithm design. Since the physical nodes of the Hadoop cluster can provide different computing resources in the case of heterogeneous distributed computing framework, which cannot achieve the best performance requirements, and by clustering the resources of different computing nodes and task scheduling resources, the communication between each task and each computing node can be reduced, to avoid the sexual bottleneck caused by heterogeneity, the computing nodes that can provide similar computing resources are considered to be the clustering of computing nodes that provide similar computing resources and the clustering of computing task Task that requires similar computing resources can be understood as combining heterogeneous physical nodes and computing tasks as closely as possible to achieve the purpose of having homogeneous clusters and homogeneous computing tasks in clusters after clustering.

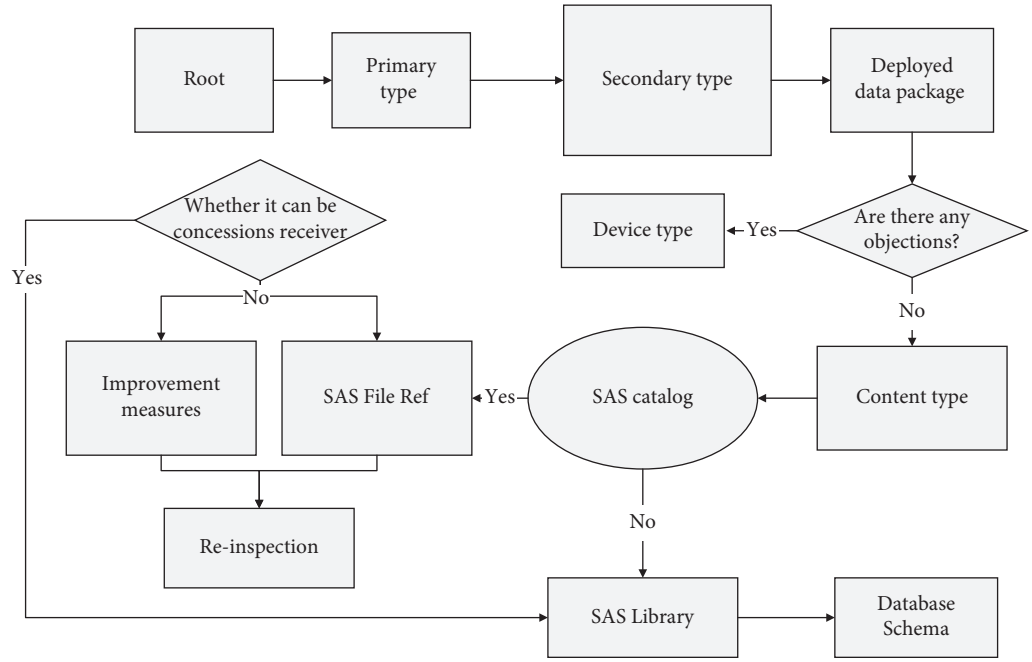

Figure 2: Schematic diagram of data locality and resource allocation.

3.3. Establishment of a Security System for Classified Computer Information Systems with the Help of Distributed Parallel Computing Programming. In the actual information security event, due to the interaction of attack and defense behaviors, the information system security attributes will change dynamically with it, and the probability of occurrence of its risk threat and the magnitude of the harm caused is also changing constantly. The traditional static model cannot correctly describe the value of information risk in the future moment and cannot judge the future moment, which greatly reduces the effect of risk assessment and makes it difficult to meet the needs of risk assessment in the confrontation process. At the same time, network attacks are covert, and to avoid the defender's tracking, attackers tend to hide some information, and information risk assessment is often incomplete for the information they obtain, but the current information risk assessment is mainly based on information system risk prediction, and the information it obtains is completely known, leading to most assessment models, and methods are not quite the same as the actual security events. The complexity of the maximum-minimum distance method iterations is based on the number of data points; however, for massive data, data types and objects are tens of thousands, and if the minimum-maximum distance method is not optimized, it is used to practice to the object data set. If the minimum-maximum distance method is not optimized and applied to the object data set, not only does it require huge computation but also takes a long time, and the required disk storage overhead will be more and more. The disk storage overhead required will also increase. Therefore, there is an urgent need for an analysis method of information system risk for network information incompleteness, which can be used to address the need for security risk assessment in confrontation and provide a basis for risk defense decisions [22].

Information security risk assessment index weights represent the importance degree of the index in the calculation process, and scientific and accurate assessment index weights can effectively improve the comprehensiveness and accuracy of information security risk assessment methods and further enhance the security guidance significance of the assessment results. Therefore, in this study, with the help of distributed parallel computing programming method, the AHP method is used to subjectively determine the weight of assessment indexes, the information entropy theory is used to objectively analyze the index weights, and then, the optimal combination of weighting method is used to comprehensively calculate the comprehensive weight of indexes.

Based on historical experience and expert recommendations, a judgment matrix is constructed for each level of judgment indicator by combining the scale benchmark table A_{ij}.

$$A_{ij} = \begin{bmatrix} 1 & \cdots & a_{i1} & \cdots & a_{m1} \\ \cdots & 1 & \cdots & \cdots & \cdots \\ a_{1j} & \cdots & 1 & \cdots & a_{mj} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{1n} & \cdots & a_{in} & \cdots & 1 \end{bmatrix}. \tag{8}$$

To determine the weight of each hierarchical index, the column vectors of the decision matrix A are first normalized, i.e.:

$$A_S = \left(a_{ij}\right)_{\mathrm{m}\times n} = \left(\frac{a_{ij}}{\sum_{i,j=1}^{m,n} a_{ij}}\right)_{\mathrm{m}\times n}. \quad (9)$$

To ensure that the indicator weights of each assessment level are consistent with the actual objective situation and to enhance the accuracy of the assessment method, a consistency test is required.

$$Q_k = \frac{\sum g_i y_{ik}^{\ 2}}{Cx}. \quad (10)$$

In this study, we use the network information system shown in Figure 3 as an example for experimental analysis. The system mainly consists of system security defense equipment, user host cluster, file server cluster, database server cluster, and user server cluster [23].

From the information security risk assessment index system shown in Figure 3, combined with the topology and functions of this example information system, the system is analyzed in the second layer of assessment indexes from four attributes: physical security environment, system operation security, information security confidentiality, and security confidentiality management, respectively. For the physical security environment, the evaluator mainly considers the security of the physical environment in which the information system is located and the external security factors of the system, such as whether it is in a place with electromagnetic protection capability. Combined with the third layer of assessment indicators, the system can be specifically divided into physical security measures, anti-interference capability, and preventive security to assess three indicators. For system operation security, combined with the information system topology, it can be seen that the defense devices to maintain the stability of system function services are mainly IDS network monitoring devices, firewalls, and anti-sniffing devices. The analysis is mainly conducted from the perspective of system security defense facilities and the main functional clusters of the system, which can be divided into four indicators for evaluation: security protection capability, data monitoring level, data storage security, and user system security. For sampling, result precision and timeliness cannot be combined, and in a large sample set, precision is better, but timeliness is poor; in a small data set, although timeliness is good, precision is poor. In big data sets, although good in timeliness, poor in precision. When the data size is larger than a certain threshold value, the precision no longer increases significantly. For information security and confidentiality, the system mainly refers to the security of information data including database information, user host information, and file server information. From the basic characteristics of information security attributes, it can be further divided into four evaluation indicators of information integrity, confidentiality, authenticity, and resistance to repudiation. The above four indicators can make an intuitive judgment on the system information security and confidentiality attributes. For security and confidentiality management, this index system is mainly evaluated from the degree of individual, organization, and management process, which can be specifically divided into whether the management system is perfect, the quality of management personnel, and the process of cryptographic key management with the information system example in this section. For the convenience of analysis, three evaluation indicators are recorded as a management system, management personnel, and password management, respectively.

Security risk assessment in the confrontation process requires both a comprehensive analysis of the risk change process under the joint action of attack and defense and research to describe the impact of information incompleteness in the actual situation on the assessment process. The traditional information security risk assessment technology mainly performs unilateral analysis from the perspective of security management and does not consider the attack method adopted by the attacker in breaking through the defense and the possible results of the network attack and defense process, thus reducing the accuracy and reliability of information system security risk assessment. On the other hand, in the process of network attack and defense, the unilateral benefits and costs are not determined by the unilateral strategy alone but are influenced by the joint decision of both parties. The problem of network information system security risk assessment has been the focus of research in the field of network security, and the application of game theory to the field of network information security has also become a current research hotspot. This chapter constructs the network attack and defense game model based on the static Bayesian game theory with incomplete information, converts the static Bayesian attack and defense game model with incomplete information into a priori judgment on the types of players, analyzes the game equilibrium, uses the pure strategy equilibrium and mixed strategy equilibrium to analyze the two situations, designs the network security risk assessment algorithm based on the static Bayesian game model, and compares it with. The effectiveness of the proposed model and method is verified through simulation experiments, which can provide an effective analysis method and model for the information security confrontation process and provide a guiding basis for the implementation of information system security emergency plan and defense strategy.

4. Experimental Verification and Conclusion

4.1. Comparison of the Standard Deviation of the Absolute Value of the Information System Security Posture Error. The information system security posture prediction model proposed in this study is validated based on the time series of the security posture values of each dimensional indicator in the criterion layer and the time series of the total system security posture values. In this paper, the 10 security posture values from the 1st assessment to the 10th assessment of the total system security posture are used as the initial security posture actual value series. The 38 actual values of the total security posture from the 11th assessment to the 48th assessment are compared with the corresponding predicted values of the system security posture to test the prediction effect of this model. The original security posture values are

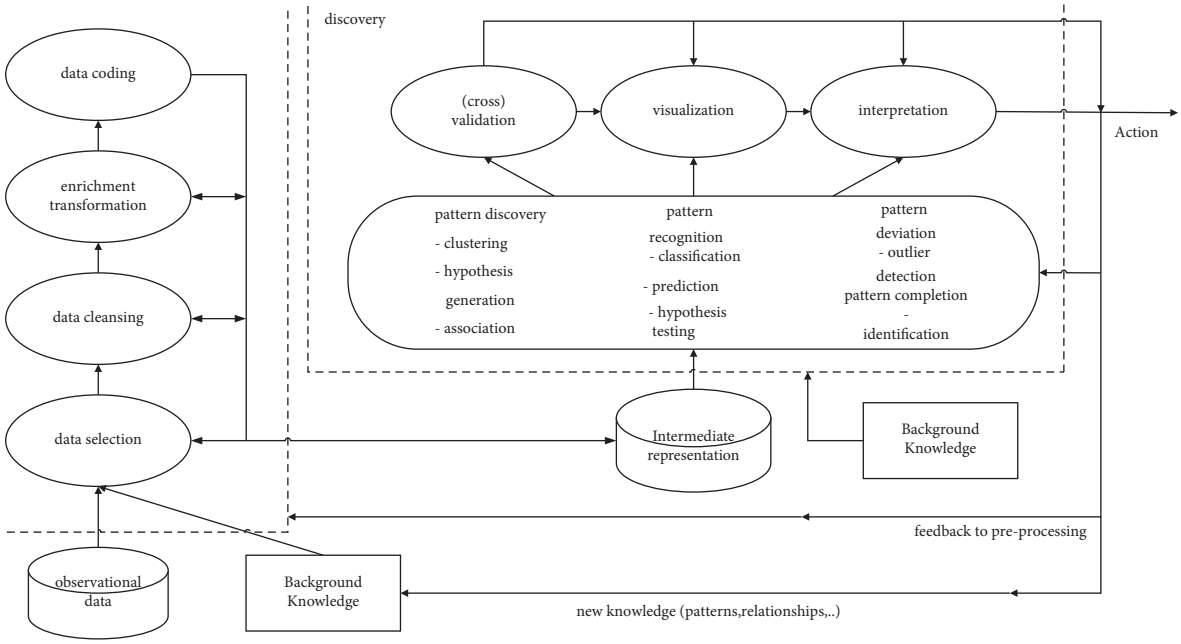

Figure 3: Cluster network information system structure.

scaled up to 10000 times the original values and normalized to the security posture values in the interval [0, 10000] to facilitate the prediction of security posture values and the classification of error states. The prediction process of the model is illustrated by taking 10 security posture values obtained from the 11th to the 20th evaluation as an example. The sequence of the 10 evaluated security postures is shown in Figure 4.

The mutation coefficient is better to reduce the error of the security posture value at T6 and T10 and improve the accuracy of the prediction results. $c = 40.0156$ is calculated as the posterior difference ratio, and the small probability test results $P = 0.89$. As can be seen from Figure 4, the prediction accuracy of the model is level 2, which satisfies the threshold test condition, then the predicted value of the security posture error of T11 is 0.04, and the mutation coefficient $\gamma = 1.834$. The predicted value of the security posture of T11 is as follows:

$$\gamma = \int k^{-i2\pi\omega x} m(z) f(x) dx = 6975.12. \quad (1)$$

The different lengths of the sequence of security posture values within different sliding window widths L will affect the values of the static smoothing coefficients in the sliding adaptive cubic exponential smoothing model and lead to a different accuracy of the initial prediction results. In this study, the sliding window width is chosen to generate a sequence of predicted values with higher accuracy before correcting the predicted values and to reduce the burden of the predicted value correction sub-module. From this study, we take $L = 5$, 10, 15, 20, and 25 and compare l with the generated initial predicted values of total information system security posture, and the mean square error of initial predicted values of each dimensional indicator security posture in the criterion layer, and the comparison of absolute value standard deviation of error is shown in Figure 5.

The smaller the mean square error, the better the initial prediction accuracy; the smaller the absolute standard deviation of the error reflects the stability of the initial error series, the smaller the absolute standard deviation of the error, the smaller the dispersion of the initial prediction and the actual value of the error, which will reduce the difficulty of dividing the subsequent error interval. In this way, it is necessary to consider the mean square error and the absolute standard deviation of the error to determine the sliding window width of the security posture prediction for different dimensional indicators. The standard deviation of the time series of the actual value of the security posture of network and data dimensions is small, and when the sliding window width is $L = 15$, the mean square error of the corresponding initial prediction sequence is the smallest; the standard deviation of the time series of the actual value of the security posture of host system and personnel management dimensions is large, and when the sliding window width is $L = 5$, the mean square error of the corresponding initial prediction sequence is the smallest; the standard deviation of the time series of the actual value of the security posture of total posture and physical environment dimensions is moderate, and when the sliding window width is $L = 5$, the mean square error of the corresponding initial prediction sequence is the smallest. The standard deviation of the time series is moderate, and when the sliding window width is $L = 10$, the mean square error of the corresponding initial predicted value series is the smallest. The reasons for this are analyzed as follows:

(1) When the standard deviation of the actual value series of security posture is small, the fluctuation of security posture is small, and the prediction effect of

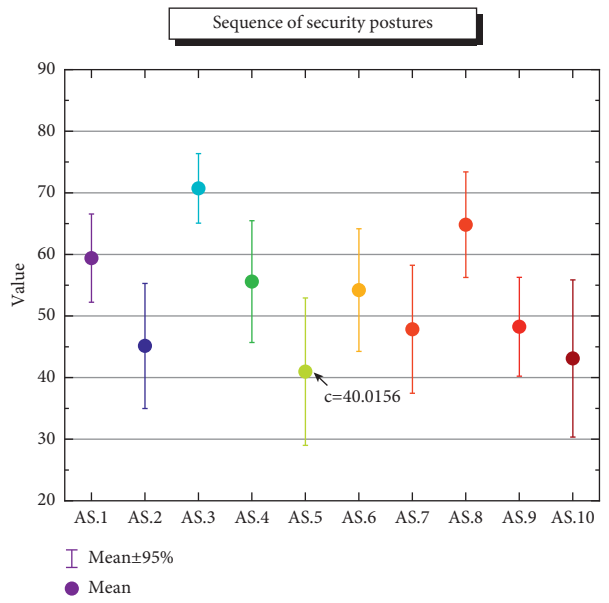

Figure 4: A sequence of security postures for the first 10 assessments.

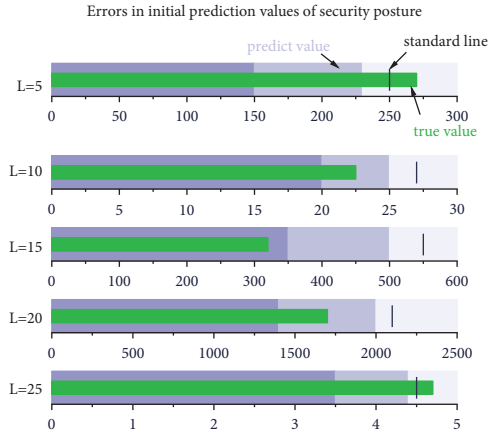

Figure 5: Comparison of errors in initial prediction values of security posture.

recent data and distant data on the value of security posture is approximately the same, so when the sliding window width is larger, the period is large, and the number of historical data contained in the window is increased, which is conducive to the calculation to obtain a more reasonable static smoothing coefficient and reduce the initial predicted and actual values errors.

(2) When the standard deviation of the actual sequence of security posture values is large, the security posture fluctuates significantly, and the prediction effect of the recent data on the security posture values is greater compared with the distant data. If the width of the sliding window is large, when one or several security posture values in the window fluctuate greatly, other less volatile data in the window will play a dominant role in the value of the static smoothing coefficient, leading to a reduction in the adaptability of the information system security posture prediction sub-module to abnormal security posture fluctuations, so reducing the width of the sliding window can divide the fine-grained security posture values. Therefore, reducing the sliding window width can divide the fine-grained security posture value sequence fragments and improve the degree of fitting between the initial predicted value sequence and the actual value fluctuation situation.

4.2. Information System Security Posture Performance Comparison. To verify the advantage of Spark distributed cluster over Hadoop in iterative computation, this experiment compares the running time of the original K-means algorithm and the optimized K-means algorithm proposed in this study on Hadoop and Spark clusters, respectively, under the same machine configuration environment and the same number of computation nodes, to reflect the respective running performance. The experimental results are shown in Figure 6.

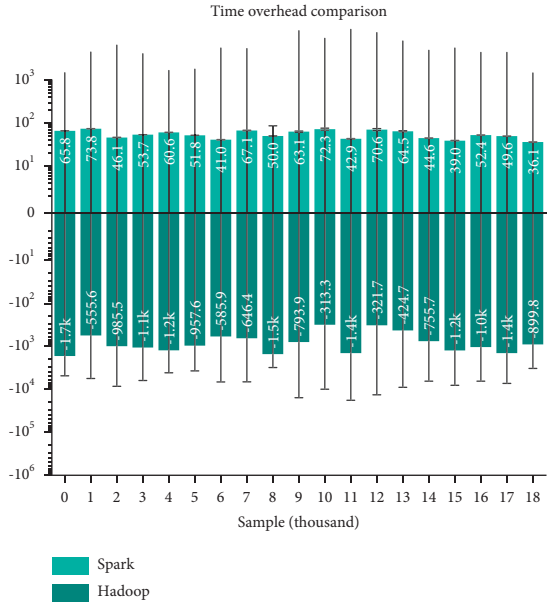

Figure 6: Time overhead comparison experiment.

This is because the Spark platform is built on top of the Yarn resource manager (Hadoop 2.x), and the advantage of Spark in-memory computing is only reflected when the data volume is huge, while the internal communication accounts for a higher proportion of the total computing time when the data volume is small. The advantage of Spark in-memory computing is realized when the volume of data is huge, but when the volume of data is small, the internal communication accounts for a high proportion of the total computing time, resulting in low data processing efficiency. In the early stage of algorithm execution, the running time of Spark is slightly higher than that of Hadoop in both the original K-means algorithm and the optimized algorithm in this study, but with the increase in data volume, starting from 15000 data, the running time of the same data volume is shorter in Spark platform compared with Hadoop platform. Spark is more efficient than Hadoop as the size of the data increases, and the trend is more obvious. This is because Hadoop reads the data object set into local memory during each iteration of computation and must reaccess HDFS, while Spark is based on in-memory computation, which significantly reduces the time overhead rate of data I/O in iterative computation, so the running time of both the original K-means algorithm and the optimized K-means algorithm in the Spark platform environment is significantly reduced compared with Hadoop platform.

In the read performance test, the same load generator used the reading test. Multiple users were simulated to read the data, where a single data size of 200 bytes was stored and read in multiple sessions. The optimization algorithm in this study performs one more traversal to find the random farthest nearest distance point, which improves the initial two initial centroids' dissimilarity and increases the accuracy of the algorithm to some extent. The read performance of the storage module is tested in two cases: random reads and range read. Figure 7 shows the average time for a single client to initiate 1000 random reads with 2 million, 5 million, 10 million, 20 million, 40 million, and 80 million data items, respectively. The horizontal coordinate represents the data size, and the vertical coordinate represents the average time spent in milliseconds.

This is because the time overhead of calculating the data storage location is the same each time, and the data range and data offset of each data block are recorded inside the storage node. Therefore, the query performance of the random query is similar at different data sizes. The performance of range queries, which are query requests with a range of start and end times per input, is illustrated in the figure. The average time to issue 10 range queries is tested for a single client with a range of 40,000 or 200,000 data items. As the amount of data in the range query gets larger, it takes more time, as expected. The size of the data store has little impact on range query performance, for the same reason as random queries, where the larger the data size, the smaller the increase in index search time.

In the total posture prediction value comparison test, based on the time series of the actual values of the security posture of information system indicators in each dimension, the prediction model of this study, the traditional Markov prediction model, and the grey Verhulst–Kalman (GVK) prediction model are used to generate the security posture prediction value sequence. Generally, the comparison of the prediction series is shown in Figure 8, where the actual value corresponding to the horizontal coordinate period $T = 200$ is the security posture value obtained from the 11th assessment in the original assessment series, and the actual value corresponding to $T = 600$ is the security posture value obtained from the 48th assessment in the original assessment series.

From Figure 8, it can be seen that the prediction model of this study generates a sequence of predicted posture values that fits the actual values better than the other two prediction models. The smaller the c value of the posterior difference

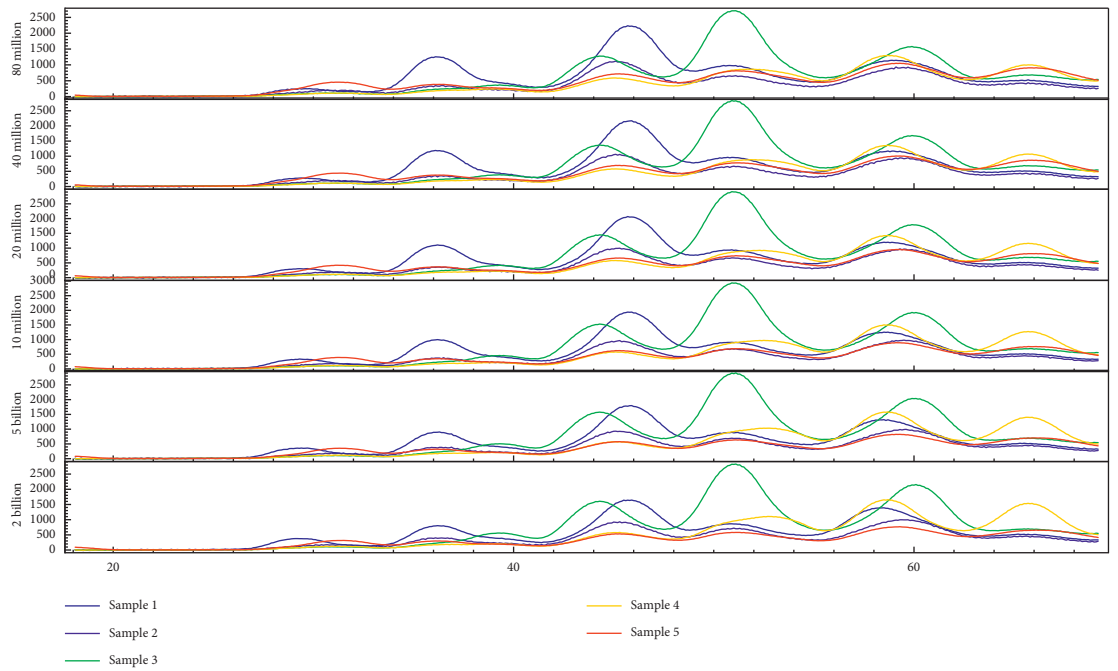

Figure 7: Read performance test results.

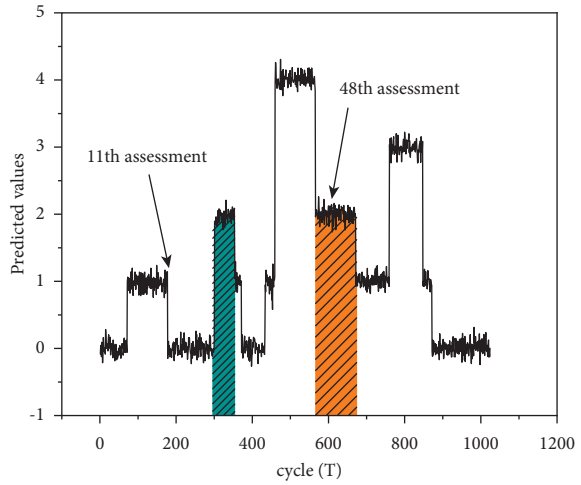

Figure 8: Comparison of the predicted values of the total situation.

test and the larger the p value of the small probability test, the better the accuracy of the prediction sequence. This indicates that the prediction model of this study is more adaptable to the prediction of variable length series. In conclusion, the prediction model in this study can improve the accuracy of information system security situation prediction.

5. Conclusion

The wide application of information systems exposes them to many risks. Information system security posture assessment and prediction are important for timely detection of potential threats to information systems and early preventive measures to reduce the possible economic losses to information systems. Therefore, this study proposes to conduct research around information system security posture assessment methods and prediction methods with the help of distributed parallel computing programming algorithms, and the specific work includes the following:

(1) Information System Security Posture Assessment Index System Research. By analyzing the information system security assessment standards, we establish the information system security posture assessment index system, design assessment indexes and assessment rules for physical environment dimension, host system dimension, network dimension, data dimension, and personnel management dimension, and provide indexes for assessing the information system security status from both qualitative and quantitative indexes.

(2) Research on Information System Security Posture Assessment Model. An information system security posture assessment model based on the interval matrix correction method is proposed. The optimal deterministic matrix is searched, and the weight vector of the index layer is determined through the modification of the interval judgment matrix, and the security posture is quantified and graded by combining with the entropy power affiliation cloud. The feasibility and validity of this paper's model are verified by evaluating the security posture of an airport departure control system (DCS). Comparing and analyzing the evaluation results with those based on entropy weight coefficient method and traditional AHP method, it shows that the security posture quantification calculation method of this study has better reliability and stability.

(3) Research on Information System Security Posture Prediction Model. An adaptive security posture prediction model based on the nonlinear time series of information system security posture assessment posture values is proposed. Based on the sliding window mechanism, the security posture values in time series are fragmented, the adaptive three-time exponential smoothing method is applied to initially generate the security posture prediction results, and the time-varying weighted Markov chain is used to predict the errors and correct the initial security posture prediction values. The experimental results show that the error between the predicted and true safety posture values obtained by the prediction model in this study is small and the fit is good. Compared with the traditional Markov prediction model and the grey Verhulst–Kalman prediction model, the prediction model in this study performs better in terms of safety posture prediction accuracy and adaptive prediction for variable time series.

Information system security posture assessment and prediction are important for timely detection of potential threats to information systems and early preventive measures to reduce the economic losses that information systems may suffer, and subsequent more optimized and efficient algorithms can be proposed from practical applications.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Acknowledgments

This work was supported by the School of Information Engineering, Henan University of Animal Husbandry and Economy.

References

[1] E. Reficco, F. Layrisse, and A. Barrios, "From donation-based NPO to social enterprise: a journey of transformation through business-model innovation," *Journal of Business Research*, vol. 125, pp. 720–732, 2021.

[2] P. Kumar, G. P. Gupta, and R. Tripathi, "A distributed ensemble design based intrusion detection system using fog computing to protect the internet of things networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 10, pp. 9555–9572, 2021.

[3] H. Faruque Aly, K. Mason, and W. Onyas, "The institutional work of a social enterprise operating in a subsistence marketplace: using the business model as a market-shaping tool," *Journal of Consumer Affairs*, vol. 55, no. 1, pp. 31–58, 2021.

[4] I. M. Ibrahim, "Task scheduling algorithms in cloud computing: a review," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 4, pp. 1041–1053, 2021.

[5] S. Tykkyläinen and P. Ritala, "Business model innovation in social enterprises: an activity system perspective," *Journal of Business Research*, vol. 125, pp. 684–697, 2021.

[6] L. Belcastro, F. Marozzo, and D. Talia, "Programming models and systems for Big Data analysis," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 34, no. 6, pp. 632–652, 2019.

[7] F. A. Goni, A. Gholamzadeh Chofreh, and Z. Estaki Orakani, "Sustainable business model: a review and framework development," *Clean Technologies and Environmental Policy*, vol. 23, no. 3, pp. 889–897, 2021.

[8] H. Shukur, S. Zeebaree, R. Zebari, O. R. Ahmed, L. M. Haji, and D. M. Abdulqader, "Cache coherence protocols in distributed systems," *Journal of Applied Science and Technology Trends*, vol. 1, no. 3, pp. 92–97, 2020.

[9] P. W. Moroz and E. N. Gamble, "Business model innovation as a window into adaptive tensions: five paths on the B Corp journey," *Journal of Business Research*, vol. 125, pp. 672–683, 2021.

[10] N. Iheanachor, Y. David-West, and I. O. Umukoro, "Business model innovation at the bottom of the pyramid–A case of mobile money agents," *Journal of Business Research*, vol. 127, pp. 96–107, 2021.

[11] H. Zhang, H. Xiao, Y. Wang, A. M. Shareef, M. S. Akram, and M. A. S. Goraya, "An integration of antecedents and outcomes of business model innovation: a meta-analytic review," *Journal of Business Research*, vol. 131, pp. 803–814, 2021.

[12] A. Kumari, R. Gupta, S. Tanwar, and N. Kumar, "Blockchain and AI amalgamation for energy cloud management: challenges, solutions, and future directions," *Journal of Parallel and Distributed Computing*, vol. 143, pp. 148–166, 2020.

[13] N. Tyagi, A. Rana, and V. Kansal, "Load distribution challenges with virtual computing[J]. Intelligent computing in engineering," *Advances in Intelligent Systems and Computing*, vol. 1125, pp. 51–56, 2020.

[14] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: a survey," *Future Generation Computer Systems*, vol. 97, pp. 219–235, 2019.

[15] Z. S. Ageed, S. R. M. Zeebaree, M. M. Sadeeq, and S. F. Kak, "A survey of data mining implementation in smart city applications," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 91–99, 2021.

[16] M. A. Omer, S. R. M. Zeebaree, M. A. M. Sadeeq, and B. W. Salim, "Efficiency of malware detection in android

system: a survey," *Asian Journal of Research in Computer Science*, vol. 7, no. 4, pp. 59–69, 2021.

[17] K. Randhawa, R. Wilden, and S. Gudergan, "How to innovate toward an ambidextrous business model? The role of dynamic capabilities and market orientation," *Journal of Business Research*, vol. 130, pp. 618–634, 2021.

[18] K. Mishra and S. Majhi, "A state-of-art on cloud load balancing algorithms," *International Journal of computing and digital systems*, vol. 9, no. 2, pp. 201–220, 2020.

[19] A. M. S. Osman, "A novel big data analytics framework for smart cities," *Future Generation Computer Systems*, vol. 91, pp. 620–633, 2019.

[20] M. Fahmideh, F. Daneshgar, F. Rabhi, and G. Beydoun, "A generic cloud migration process model," *European Journal of Information Systems*, vol. 28, no. 3, pp. 233–255, 2019.

[21] T. Haaker, P. T. M. Ly, N. Nguyen-Thanh, and H. T. H. Nguyen, "Business model innovation through the application of the Internet-of-Things: a comparative analysis," *Journal of Business Research*, vol. 126, pp. 126–136, 2021.

[22] A. Mohamed, M. K. Najafabadi, Y. B. Wah, E. A. K. Zaman, and R. Maskat, "The state of the art and taxonomy of big data analytics: view from new big data framework," *Artificial Intelligence Review*, vol. 53, no. 2, pp. 989–1037, 2020.

[23] J. R. Gamble, E. Clinton, and V. Díaz-Moriana, "Broadening the business model construct: exploring how family-owned SMEs co-create value with external stakeholders," *Journal of Business Research*, vol. 130, pp. 646–657, 2021.