Edge-Fog-Cloud Secure Storage with Deep-Learning-Assisted Digital Twins

Zhihan Lv and Ranran Lou

Abstract

The present work studies the storage security of edge-fog-cloud computing to improve cloud storage security. The data of an intelligent manufacturing industrial machine is used in the research and pre-processed. Machine manufacturing perception data includes data storage and data transmission. In addition, combined with the perception data of machine manufacturing, digital twins technology is used to construct the digital twin in the real world to simulate the online data-driven behavior of machine manufacturing. The digital twins model is built and saved by 3Dmax. Finally, deep learning technology is introduced to defend against network intrusion. Cloud databases are vulnerable to external attacks and internal attacks of cloud data providers. Thus, the homomorphic encryption algorithm and secure multi-party computing are introduced to ensure that the database stores ciphertext data and performs data queries directly. The experimental results indicate that digital twins technology has a good effect. The two-layer cloud database model costs US$11,476.1, the lowest among comparative models. The unencrypted Amazon Relational Database Service has the best performance. The model proposed here also achieves relatively high availability, and the hybrid cloud structure significantly improves the model's performance. The research content provides a reference for realizing data storage security in a hybrid cloud.

Introduction

Cloud computing is mainly deployed in large data centers on the Internet [1, 2]. With the rapid development of computer technology, cloud computing technology has developed into a mixture of various technologies, such as utility computing, distributed computing, parallel computing, load balancing, hot backup jumbling, virtualization, and network storage [3, 4]. Traditionally, data of computers is stored in mobile hard disks or local hard disks. Once the hard disks are damaged or lost, information and data are easily leaked, with low data storage security [5, 6]. Fog computing and edge computing are extended computing technologies based on cloud computing. Fog computing is primarily oriented to the Internet of Things (IoT), and computing resources are scattered at the edge of the network in the form of miniature local data centers. Fog computing provides processing capacity within the local area network [7, 8]. Edge computing also belongs to a computing resource mode. It is deployed at the data source or in proximity to provide near-end services [9].

In practical applications, these computing methods are sometimes used alone, sometimes in pairs, and sometimes in combination to obtain obvious advantages in specific environments and occasions. For example, in the prevailing smart cities, cloud computing can provide massive data access and high-intensity data processing functions suitable for such non-real-time application scenarios as intelligent transportation. The traffic signal controller is a sensor in the receiving area of fog computing nodes to collect the flow information and traffic congestion situation and dynamically adjust signal lights' phase and time allocation after real-time analysis. At the same time, the cloud computing center analyzes the traffic situation and traffic demand of a broader region and even the whole city to improve the overall traffic environment, such as unique service traffic control, green wave control, and traffic dredging. Autonomous driving technology realizes the car's autonomous driving by working with artificial intelligence, machine vision, radar, positioning system, and automatic control systems [10].

Zhihan Lv is with the Department of Game Design, Faculty of Arts, Uppsala University, Sweden

Ranran Lou is with Qingdao University, China.

Digital Object Identifier: 10.1109/IOTM.002.2100145

In recent years, deep learning technology has also made particular development in network security. Deep learning technology has a promising application prospect in building detection models to deal with network intrusion. As the digitalization of the natural world, digital twins connect objects in the real world with themselves under the vision of IoT and synchronously provide state information and change information of things [11]. This article combines edge computing, fog computing, and cloud computing for safe network storage. First, hybrid cloud security storage and deep learning technology are introduced. Then digital twins technology is used to build the virtual model to get the machine-made database. Second, deep learning technology sets up the first intrusion detection. Finally, the data encryption technology sets up the second guarantee for secure storage. The innovation of this article lies in the edge-fog-cloud computing mode for secure data storage. The research conclusion provides a new approach for the safe storage of network data under the latest technology.

Research Status of Cloud Computing Secure Storage

After cloud computing was proposed at the Search Engine Conference in 2006, cloud storage also began to enter the public's field of vision. China's comprehensive standardization technology system for cloud computing was proposed in 2014 [12]. The cloud computing mode can provide convenient, usable, on-demand network access, and enter a configurable resource sharing pool, and the resources in the shared pool can be quickly extracted at low cost. With the rapid development of cloud computing, domestic and foreign cloud computing platforms and cloud computing service providers have gradually increased. Cloud computing service platforms include Microsoft's Azure, Amazon's Amazon Web Service, Google's Google Compute Engine, and so on. The largest cloud computing service platform is Alibaba Cloud of Alibaba Group in China.

Users of cloud storage systems do not have the right to control data management, and cloud service provider servers are vulnerable to malicious attacks, causing data leakage [13]. Previous studies have analyzed data storage security and found that data and user information are encrypted to ensure cloud storage confidentiality effectively. Research on edge computing mainly focuses on the Internet of Vehicles and other fields, which often rely on the collected data to control the equipment in real time.

 2576-3180/22/$25.00 © 2022 IEEE

Wrong control commands issued by the system due to data attacks will cause property loss and even harm personal safety. Therefore, data security research is a pain point to be solved urgently in the field of edge computing. Data security is always an important issue to ensure enterprise and government information security to perfectly integrate cloud computing and edge computing and give full play to the characteristics of device-cloud collaboration, which is worthy of systematic research.

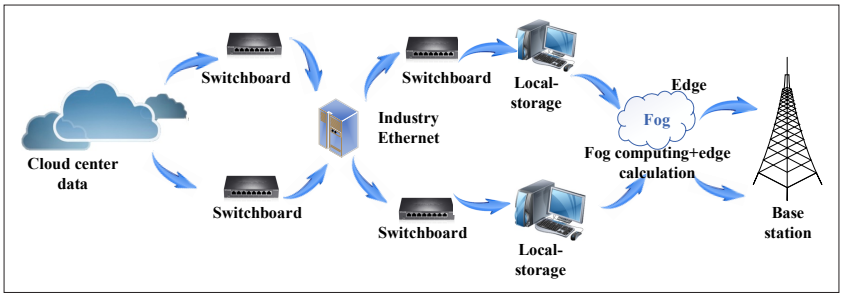

FIGURE 1. Machine data acquisition of edge-fog-cloud computing.

Previous studies focused on the theories of cloud computing secure storage methods. There were few studies on the application of cloud computing, fog computing, and edge computing to secure storage in IoT. The era of the Internet of Everything has come, and the application research of cloud computing secure storage in the IoT system is very important. Therefore, in the next section of this article, we use digital twins technology to build a combination of virtual and real, use deep learning technology to build a network intrusion prevention system, and finally build a secure storage model based on the edge-fog-cloud computing network itself.

Construction of a Network Intrusion Prevention System Based on Digital Twins and Deep Learning

Edge-Fog-Cloud Collaborative Data Acquisition

In order to combine the actual research, this article chooses an industrial machine in intelligent manufacturing as the research object. The data acquisition object of this study is machine manufacturing perception data. The machine manufacturing perception data in this article includes data storage and data transmission. The data transmission core of this article is the switch and Ethernet. The perception information is transmitted from the Zigbee base station to the 4G broadband base station to realize the construction of the overall wireless network system of the work surface, and the broadband wireless network and narrowband convergence of wireless networks. The specific structure is shown in Fig. 1.

It can be seen from Fig. 1 that the data storage in this article adopts edge-fog-cloud computing collaborative storage, and the storage methods are centralized cloud storage, distributed fog storage, and edge storage. The arrow in Fig. 1 refers to the transmission direction of machine data. In this article, wireless routers, data acquisition terminals, video monitoring terminals, and other devices are used to realize the effective transmission of pictures, videos, and data. Data storage here combines distributed edge storage, fog storage, and centralized cloud storage. The introduction of edge storage and fog storage can overcome the shortcomings of cloud storage such as low security, low bandwidth, and high latency. Edge computing and fog computing can realize local storage of data, alleviate the data analysis load of cloud center equipment, and improve the safe storage of data. Data that needs to be processed centrally and time-insensitive data will be transmitted to the cloud center, which can access edge computing and fog computing data at any time.

Digital Twins Technology Build Intelligent Manufacturing Digital Twin

First, we use digital twins technology to build the digital twin of the machine. Digital twins combine virtual analog simulation technology and digital technology to digitally process the running state of physical space objects, completing the interaction between virtual space and physical space. This research uses digital twins technology to complete the behavioral simulation of online data-driven machine manufacturing, realize the visualization of the real-time state of machine manufacturing, and provide the state of machine manufacturing under normal data processing. The construction of the machine-made digital twin is shown in Fig. 2.

As can be seen from the above figure, the construction of the digital twin model in this article is based on the physical parameters of multiple machine manufacturing and the processing parameters of the machine. In Fig. 2, the red arrow represents the data transmission direction of machine data, and the blue arrow represents the data transmission from physical space to virtual space. The model is constructed and saved in 3Dmax. The simulation environment is built in Unity3D by adopting offline data, such as location and environmental parameters. Finally, the model is converted into a format that can be recognized by Unity3D in 3Dmax and imported into Unity3D. In this way, the simulation environment and high-fidelity model are integrated. The digital model and virtual space environment can be updated in real time by running the online data and dynamic parameters to drive the high-fidelity model in the super-realistic simulation environment for high-fidelity simulation.

Deep Learning Technology to Build an Intrusion Prevention System

An intrusion prevention system is completed by intrusion detection and intrusion prevention together. The clustering algorithm needs to divide similar parts of the data sample into the same category. Thus, it assumes large enough differences between classes and high similarities in the same type to facilitate classification. Then the clustering algorithm can divide similar parts in the sample into a variety. Deep learning methods rely on their own structure to extract features from multi-dimensional intrusion data.

Sparse autoencoders make use of the mutually exclusive property of neurons to reduce the characteristic dimension of the input layer by mimicking the way humans think and make the problem sparse by representing the problem with fewer implicit units. The deep learning network based on autoencoders needs multiple feature extraction and hidden layers to obtain the classification vectors. The selection of the classifier is based on the elements of the last hidden layer. In the process, the features of the input vector are extracted through multiple hidden layers. The sparse autoencoder-based intrusion prevention neural network has a multi-layer structure. After progressive feature extraction, the result becomes more abstract. The final classification result can be obtained by classifying the abstract feature, taken as the result of intrusion detection. Based on the mutual exclusion of neurons, the sparse autoencoder network based on deep learning is selected to obtain classification vectors.

The final intrusion detection system is composed of data preprocessing, input layer, sparse autoencoder neural network layer, SoftMax classifier, result sorting, and other parts. The classification result obtained by feature extraction in this part is the result of intrusion detection. The feature extraction in this article includes the initialization of the neural network, the processing of hidden layer errors, the calculation of the weight gradient and the bias value gradient corresponding to the neural net-

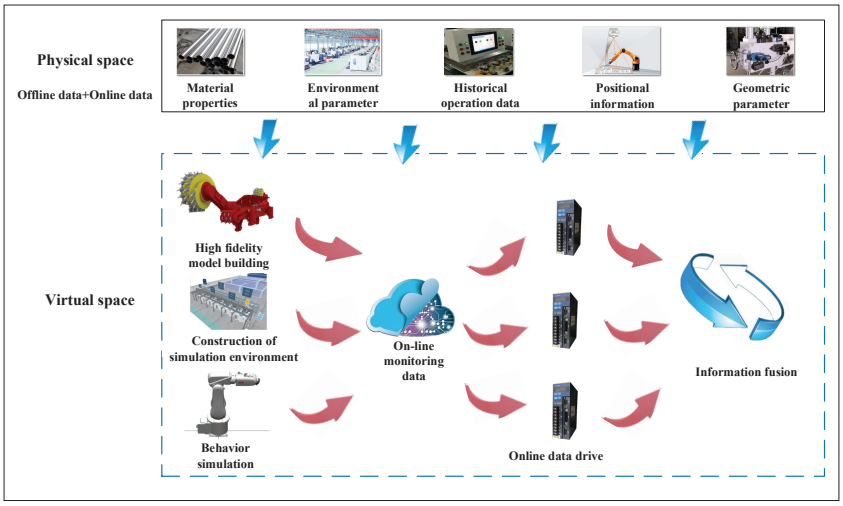

FIGURE 2. Digital twins model.

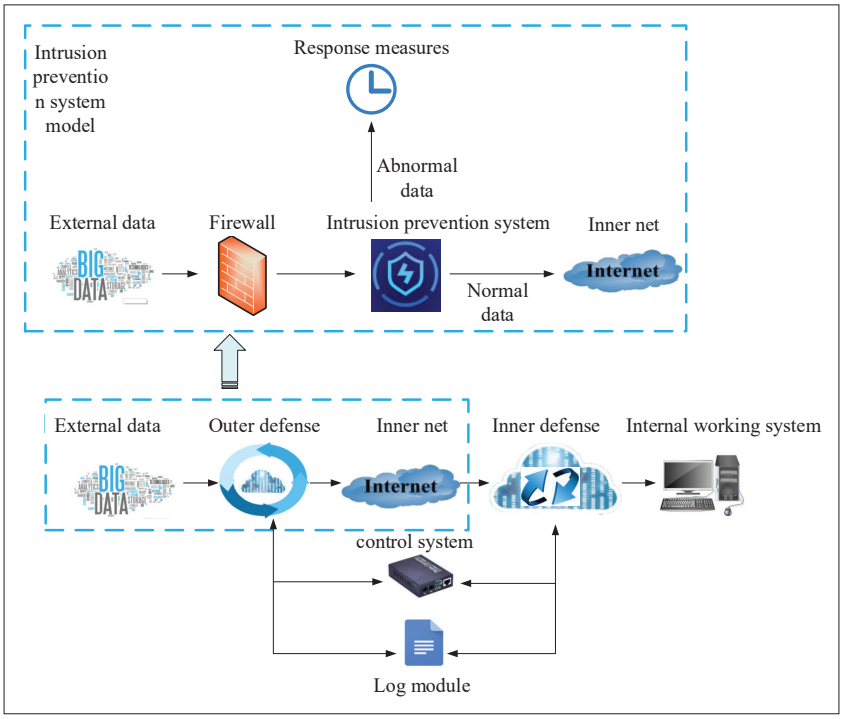

FIGURE 3. Network defense system design architecture diagram.

work, and the final update of the fine-tuned weight and bias value. Part of the data obtained in the previous section is used as the training set. The deep model intrusion detection algorithm is trained, the hidden layer error is calculated, the weight matrix and the deviation vector are adjusted, the learning result is obtained by inputting the test set, and the classification result is obtained by inputting the classification set. Finally, the false alarm rate and detection rate are obtained.

The design of the intrusion prevention system chooses a layered architecture, and the average load of the control system is within an acceptable range. The specific design structure is shown in Fig. 3. By checking external communication data, it allows normal data to enter the internal network for interaction through the firewall and blocks abnormal data to ensure that the network is not threatened. External input to the firewall will trigger the intrusion prevention system. The intrusion prevention system also checks the data type. If the data is normal, it will enter the internal network. Otherwise, the system will take response measures. The defense system consists of control, communication, intrusion prevention, and logging.

Construction of the Secure Storage Model under the Hybrid Edge-Fog-Cloud Computing

At present, there are two main homomorphic encryption approaches: fully homomorphic encryption and somewhat homomorphic encryption. Fully homomorphic encryption allows any operation of operators to be performed on the ciphertext. However, the existing fully homomorphic encryption algorithms require a large amount of computation, resulting in a limited range of applications. First, the data storage is encrypted. Based on the small amount of calculation and the wide range of application, this article selects the Rivest-Shamir-Adleman (RSA) algorithm in homomorphic encryption. The encrypted data is stored in two parts. The two parts of data are separately stored at the cloud service provider and the data owner. In this model, the data owner needs less storage space, which reduces the burden of local storage. The database system consists of users, cloud databases, and data owners. The specific structure is shown in Fig. 4.

It can be seen from Fig. 4 that there is some key data of the client in the storage data model. The communication module of the client guarantees the communication of each central data bank (CDB). When the user initiates a query request, the Structured Query Language (SQL) will analyze the request, and classify it into write query and read-only query. The SQL distribution module then sends the query request to the corresponding CDB for processing. The entire client database is transparent to the user, and user operation of the traditional database management system (DBMS) and operating client database applications are the same.

To establish a ciphertext data storage model, first create a database on the client, select the RSA algorithm, select two large prime numbers, get their product, generate a positive integer, create a new table and field column in the CDB database, and create a column key. Set the definition for the number of each row and store it independently. The storage space needs to be re-encrypted with a homomorphic encryption algorithm that supports addition. Finally, a two-column encryption algorithm table is obtained. The client only needs to store large prime numbers and column keys. The actual data of the table is stored in the CDB database. After the above steps are completed, insert data into the table. All data values must be values in the integer domain. For the plaintext data to be inserted, it needs to be encrypted by the column key and row number definition. The model uses a lot of modular operations in encryption and decryption operations, which wastes computing resources.

In the high-security available database system model, cloud servers of two or more different cloud service operators are generally used for system deployment to ensure data integrity and high availability of the database. The massive data in this model are stored in layers, so edge computing is used to assist the data analysis process of the sensor network. This article applies the model to a distributed camera sensor network to construct an intelligent monitoring system based on edge computing. The edge nodes in the edge layer can receive raw video data coming from the camera. The data is preprocessed to extract the characteristics of the target object. Then the data is uploaded to the fog server and cloud service layer. In this way, the massive video

data collected from the large-scale surveillance network can be processed in real time, while the cloud service layer only needs to focus on the integration and analysis of feature data. Due to the principle of load balancing, the load of multiple independent servers can be balanced, and the load of each independent server will be minimized. A multi-cloud database with a high-security available database system saves more costs than a single cloud database. Since this article uses the edge-fog-cloud collaborative structure, a high-security database model for a single-machine system and multiple cloud service databases is set up.

Hybrid Cloud High-Security Database Model Simulation

Finally, a simulation experiment is carried out on the database model created in this article. It is assumed here that the user needs a database with 1 transaction per second (TPS) performance, the server combination chooses an 8-core central processing unit (CPU) and 24 GB memory, and the bandwidth is 80 Mb/s. The database information in this article is the data obtained in the intelligent machine manufacturing mentioned above. The high-security available database in this article is installed on an ordinary computer, written in C language, and the client host is configured with 4 GB of memory and Core i3-3240 CPU.

First, compare and analyze the cost of different multi-cloud database models. The cost items are the annual cost of RAM, CPU, hard disk, and broadband of the standalone system, double-layer cloud system, three-layer cloud system, and four-layer cloud system. The cost comparison of different solutions will determine the practicality of the model created in this article. Second, the next section compares the security storage performance of different models. Among them, Amazon's RDS model is an unencrypted plaintext system, and the other models are all encrypted systems. The experiment running time is 30 min, and the available performance of the model is characterized by the non-timeout rate of data query, and the timeout standard is 5000 ms. Since there are fewer models for secure storage, and the core part of the model comparison is to compare the performance of encrypted databases, the comparison databases used include Microsoft, Alibaba Cloud, Tencent, and Amazon cloud databases as the CDB side.

Data Storage Security Experiment Analysis and Performance Evaluation

Cost Analysis Results of Different Multi-Cloud Database Models

The cost of different multi-cloud database solutions is compared below, and the results are shown in Table 1.

Table 1 indicates that the total cost of the two-tier cloud database model is US$11,476.1, which is the lowest compared to US$11,830.8 for the standalone model, US$12,291.4 for the three-tier cloud database model, and US$13,282.4 for the four-tier cloud database model. It shows that users can use the model of this article to complete the deployment of encrypted databases on the premise of choosing a cloud service provider, without adding additional costs. The three-tier cloud database model and the four-tier cloud database model are suitable for more demanding scenarios and can avoid the problem of operator blockade. The cost is 4 and 12 percent more expensive than the standalone model.

Model's Secure Storage Performance Verification

After the first defense system, the availability comparison of the encrypted database of this article and the unencrypted database is shown in Fig. 5.

It can be seen from Fig. 5 that within the time range set in this experiment, less than 100 percent of the queries can be completed. Amazon RDS, an unencrypted database, has the best performance with system throughput of 1.1 TPS, which is basically the same as the set value. The performance data of the encrypted database is increased by the encryption operation, so the query operation time of the database is longer than that of the unencrypted database, and the query success rate and system throughput are also weaker than the unencrypted performance. However, in an encrypted database, the performance of the model in this article is stronger than that of other databases, and the performance of the database in this article obviously depends on the hierarchical data model and multi-cloud server. The more databases used, the higher the success rate of the model, which means that the model proposed in this article has high availability, and the hybrid cloud structure significantly improves the performance of the model.

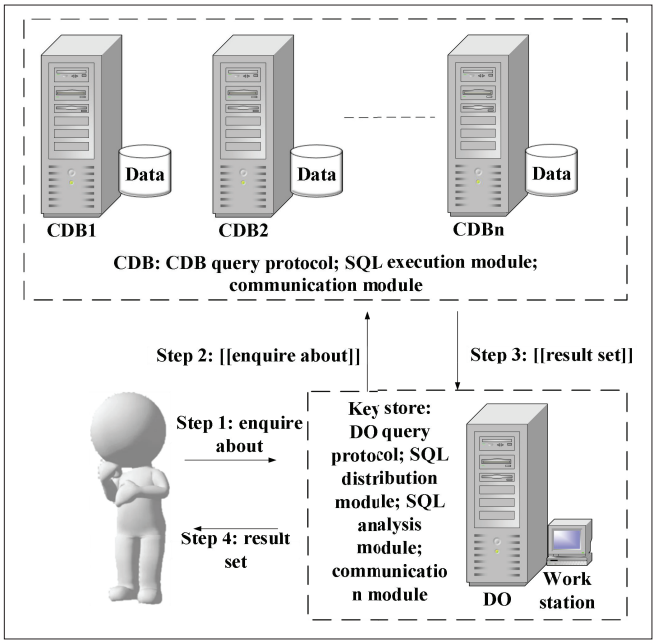

FIGURE 4. Network defense system design architecture diagram.

	Standalone system	Two-tier cloud system	Three-tier cloud system	Four-tier cloud system
RAM(GB)	24	12	8	6
CPU (number of cores)	8	4	2	2
Hard disk (TB)	1	1	1	1
Bandwidth (Mb/s)	80	42	28	24
Frequency doubling coefficient	5. 0	4. 0	2. 0	2. 0
Server front-end bus frequency (GB/s)	6. 4	3. 2	1. 8	1. 8
Processor external frequency (MHz)	200	200	200	200
Total cost (year/dollar)	11,830.8	11,476.1	12,291.4	13,282.4

TABLE 1. Cloud service cost comparison of multi-cloud database solutions.

Discussion

The cost analysis of the database model based on edge-cloud-fog computing reported here suggests that the cost of the two-layer cloud database model is US$11,476.1, which is lower than the model constructed by Tos *et al.* (2021) [14]. The security storage performance verification shows that the hybrid cloud structure significantly improves the model's performance; the more the database is used, the higher the model's success rate. Therefore, the model constructed here has a promising application prospect. It is consistent with Ambica (2021) research results, indicating that hybrid cloud structure plays a positive role in improving model performance [15].

Conclusion

Considering the high risk of cloud data storage, this article introduces the research background of secure storage, combining digital twins technology, deep learning, and edge-fog-cloud computing to study storage security. Deep learning is the first step to intercept malicious intrusion in network security storage, and edge-fog-cloud computing and encryption technology make up the second step to guarantee data security storage. The experimental results demonstrate that the cost of the model proposed here is US$11,476.1, which is relatively low. The model also achieves excellent performance; in particular, the system throughput of the unencrypted Amazon's RDS model is 1.1104 TPS. There are still some deficiencies in this article. The simulation experiment only compares the cost difference under different databases but does not compare the edge-fog-cloud algorithm with other algorithms. Hence, the follow-up study will compare the edge-fog-cloud algorithm and other algorithms, and add a quantitative evaluation of the step-by-step defense and the overall defense system to make the results more convincing.

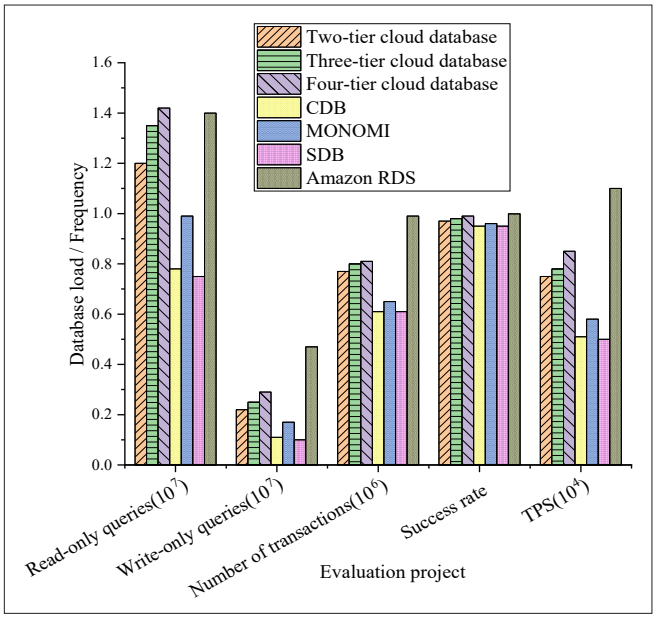

FIGURE 5. Usability evaluation results of different database models. Note: The unit of success rate is percent.

References

[1] M. Paprzycki, "Towards Edge-Fog-Cloud Continuum," *Procedia Computer Science*, vol. 179, 2021, p. 3.

[2] M. Aslam *et al.*, "FoNAC-An Automated Fog Node Audit and Certification Scheme," *Computers & Security*, vol. 93, 2020, p. 101,759.

[3] P. Kochovski *et al.*, "Trust Management in a Blockchain Based Fog Computing Platform with Trustless Smart Oracles," *Future Generation Computer Systems*, vol. 101, 2019, pp. 747–59.

[4] G. Merlino *et al.*, "Enabling Workload Engineering in Edge, Fog, and Cloud Computing Through OpenStack-Based Middleware," *ACM Trans. Internet Technology*, vol. 19, no. 2, 2019, pp. 1–22.

[5] J. Feng, L. T. Yang, and R. Zhang, "Practical Privacy-Preserving Highorder Bi-Lanczos in Integrated Edge-Fog-Cloud Architecture for Cyberphysical-Social Systems," *ACM Trans. Internet Technology*, vol. 19, no. 2, 2019, pp. 1–18.

[6] M. Jayaram and H. Fleyeh, "Whither Edge Computing? A Futuristic Review," *Int'l. J. Applied Research Info. Technology and Computing*, vol. 9, no. 2, 2018, pp. 180–88.

[7] R. Soman and R. Sukumar, "Secure Storage and Sharing of Visitor Images Generated by Smart Entrance on Public Cloud," *Int'l. J. Digital Crime and Forensics*, vol. 13, no. 5, 2021, pp. 65–77.

[8] S. E. Ebinazer, N. Savarimuthu, and S. M. S. Bhanu, "ESKEA: Enhanced Symmetric Key Encryption Algorithm Based Secure Data Storage in Cloud Networks with Data Deduplication," *Wireless Personal Commun.*, vol. 117, no. 4, 2021, pp. 3309–25.

[9] J. Tian, H. Wang, and M. Wang, "Data Integrity Auditing for Secure Cloud Storage Using User Behavior Prediction," *Computers & Security*, vol. 105, 2021, p. 102,245.

[10] R. Williams, J. A. Erkoyuncu, and T. Masood, "Augmented Reality Assisted Calibration of Digital Twins of Mobile Robots," IFACPapersOnLine, vol. 53, no. 3, 2020, pp. 203–08.

[11] P. Franciosa *et al.*, "Deep Learning Enhanced Digital Twin for Closed-Loop In-Process Quality Improvement," *CIRP Annals*, vol. 69, no. 1, 2020, pp. 369–72.

[12] J. Lee *et al.*, "Integration of Digital Twin and Deep Learning in Cyber-Physical Systems: Towards Smart Manufacturing," *IET Collaborative Intelligent Manufacturing*, vol. 2, no. 1, 2020, pp. 34–36.

[13] K. Arafet and R. Berlanga, "Digital Twins in Solar Farms: An Approach through Time Series and Deep Learning," *Algorithms*, vol. 14, no. 5, 2021, p. 156.

[14] U. Tos *et al.*, "Achieving Query Performance in the Cloud via a Cost-Effective Data Replication Strategy," *Soft Computing*, vol. 25, no. 7, 2021, pp. 5437–54.

[15] V. Ambica, "Hybrid Cloud Security Measures and Research Challenges," *Turkish J. Computer and Mathematics Education*, vol. 12, no. 10, 2021, pp. 3578–85.

Biographies

Zhihan Lv [SM'19] is a senior lecturer/associate professor at Uppsala University, Sweden. He obtained his Ph.D. degree from Ocean University of China in 2012. His research interests include digital twins, the Internet of Things, virtual reality, and blockchain. He has published 300+ papers in top journals and conferences.

Ranran Lou received his Bachelor's degree in engineering from Qingdao University in 2019, where he is currently pursuing a Master's degree in software engineering with the School of Data Science and Software Engineering. He has extensive experience in marine data processing. His research directions are deep learning and big data.