
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 5, pp. 3484–3488, DOI 10.12694/scpe.v25i5.3034

DATA PROTECTION AND PRIVACY PROTECTION OF ADVERTISING BASED ON CLOUD COMPUTING PLATFORM

ZHISHE CHEN*

Abstract. This paper uses the hybrid leapfrog algorithm to mine user information in encrypted advertisements effectively and intelligently. This method handles the nonlinearity of the original data by mapping it into kernel space. The representation of the original ciphertext in the kernel space is obtained by sparsely reconstructing the encrypted original advertising data. Build a corresponding scoring mechanism and select the best advertising data characteristics. The selected data were clustered using the data fuzzy clustering method based on the improved hybrid leapfrog. Set the adjustment coefficient to improve the local optimization performance of hybrid frog leaping. This algorithm uses the tightness and separation in genetic algorithms and constructs a fitness function to determine the clustering critical value. This enables the practical, intelligent mining of homomorphic passwords with privacy protection. Experimental results show that the method proposed in this article can effectively improve the convergence speed and accuracy of clustering. Improve Blowfish by combining multi-threading, sharing encryption and other methods. This enables encryption and decryption of large amounts of model data. The research of this project has very important research value in improving the security performance and effectiveness of cryptographic algorithms.

Key words: Privacy-preserving deep learning; Precise advertising; Secure computing; Privacy protection; Data mining

1. Introduction. Advertising data mining technology has become an emerging research direction. Advertising data mining technology has been widely used in many fields, such as engineering, scientific research, etc. This provides users with revenue. But when this method is used for ad mining, it can also have adverse effects, like the security of personal information. Technicians can speculate that it may contain private or sensitive information. There is a potential risk of leakage when mining advertising data with homomorphic passwords, so it must be protected. It is necessary to conduct data mining while ensuring the security of ciphertext information. Especially with the rapid development of network technology, scientific researchers can obtain massive amounts of information from various web pages. Traditional advertising user data mining methods to mine privacy protection data based on ciphertext may cause data leakage and affect the accuracy of data mining. Literature [1] uses homomorphic cryptography technology on the lattice to mine data. It implements data mining for privacy-preserving data cluster analysis. The user encrypts the data before transmitting it to the provider. After using Blowfish-based confidentiality and data mining technology, network service providers cannot obtain user data. Compared with existing data publishing methods, Geji data mining technology has significant advantages in information security. This algorithm ensures the accuracy of the spacing of integer encrypted ciphertext. Experiments have shown that the computing speed of the lattice mining algorithm is significantly higher than that of other algorithms, but there are also problems with low accuracy. Literature [2] proposed a time series data mining algorithm for differential privacy—screening of sequential patterns from candidate templates. Geometric principles are used to add noise to selected mode support values to interfere with their generation. The simulation results show that the proposed algorithm meets the differential confidentiality requirements. Although it has a good mining effect, the accuracy is not high [3]. This paper proposes homomorphic encryption data privacy protection technology and model privacy protection technology based on the Blowfish algorithm.

2. Homomorphic encryption technology. Homomorphic encryption is an encryption method based on computational complexity. Perform corresponding operations on the homomorphically encrypted data and then decode it. A similar conclusion was made for the unencrypted raw material [4]. Homomorphic encryption is a

*School of Information Science and Engineering, Wuchang Shouyi University, Wuhan, Hubei, 430064, China (Corresponding author, czs_works@163.com)

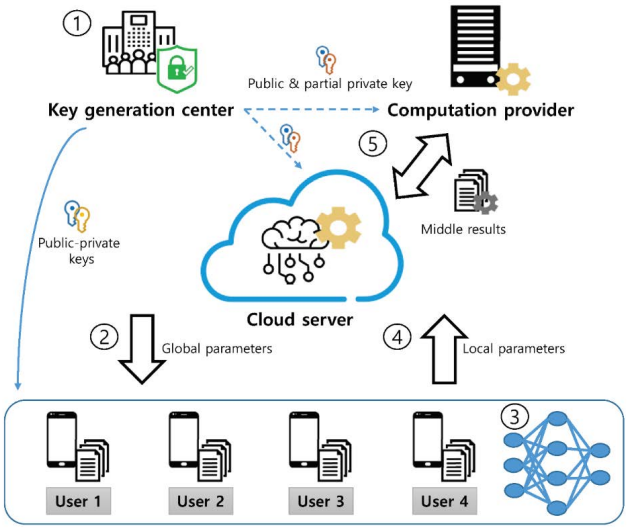

Fig. 2.1: *Homomorphic encryption intelligent reasoning architecture.*

new computing method that has emerged in recent years. It can effectively ensure the security of information transmission and calculation [5]. Homomorphic ciphers can be used to encode ciphertext directly. Ciphertext calculation eliminates the cumbersome process of decrypting large-scale ciphertext, then calculating, and then encrypting. This algorithm saves the amount of calculation while ensuring safety. This method can provide ideas for research on privacy protection issues of advertising users in the network environment. Any individual server can hide trained models. These models can be secretly exchanged between multiple servers to ensure model security. Figure 2.1 shows the intelligent inference architecture based on homomorphic cryptography.

The implementation process of homomorphic encryption advertising user data privacy protection technology for intelligent algorithms is as follows: (1) Use methods such as homomorphic addition and homomorphic multiplication to establish convolution and pool and approximate different excitation functions; (2) Based on security, Verify the security of ciphertext through comprehensive multi-party computation; (3) Study the security and privacy protection mechanism based on ciphertext. The above process applies to 6 different machine learning algorithm scenarios: convolutional neural network, BP neural network, logistic regression algorithm, SVM, linear regression algorithm and multi-layer perceptron [6]. The inference engine consists of three servers, which can automatically encrypt user information locally without decrypting it. It only needs to perform cryptographic operations on it and then feedback the results of its operations to the user. The user decrypts it to obtain inferred conclusions [7]. The user has no sense of the above inference. Only the user can encrypt it, ensuring the security of the entire computing process. This paper studies cryptographic intelligence algorithms based on Parlier's cryptography ideas. The algorithm of Parlier homomorphic cipher is:

$$z = \mathrm{Enc}(u, s) = h^u s^m \bmod m^2$$

h and m are the password public keys; $s \in M.s$ and m are mutually prime. h is a random number much smaller than m^2, and is generally more than 4000 digits, so its operation cost is very huge [8]. The calculation formula of Partier homomorphic addition is:

$$\mathrm{Enc}(u, s)\,\mathrm{Enc}(v, \varepsilon) = \left(h^u s^m \bmod m^2\right)\left(h^v \varepsilon^m \bmod m^2\right) =$$
$$h^{(u+v)}(s\varepsilon)^m \bmod m^2\,\mathrm{Enc}(u+v, s\varepsilon)$$

ε is the same random number as $s.s$ is the random number for pairing u, and ε is the random number for pairing v. The formula for Parlier homomorphic multiplication is:

$$\mathrm{Enc}(u, s)^\kappa = \left(h^u s^m \bmod m^2\right)^\kappa = h^{u\kappa}\left(s^\kappa\right)^m \bmod m^2 = \mathrm{Enc}\left(u\kappa, s^\kappa\right)$$

κ is a sub-square number. Information was exchanged based on confidential sharing. The confidential multi-party computation method is used in this paper [9]. This algorithm allows multiple parties to jointly keep their value secret without destroying computing power [10]. This method divides a piece of data into several segments and does not display the original data when sharing. Two operation participants perform the same operation on an encryption key group and then reassemble it. On the user side, the private information u is decomposed into two parts, u_0 and $u_1, u = u_0 + u_1$. Then the two variables u_0 and u_1 are sent to the two servers D_0 and D_1 respectively [11]. Mere possession of u_0 and u_1 does not jeopardize the privacy of data u. The formula for decomposing tensor u is this:

$$u_0 = \text{share}(u, S) = S \bmod N$$
$$u_1 = \text{share}\,2(u, S) = u - S \bmod N$$

S and N are both arbitrary numbers. Two servers Q_0 and Q_1 and one secondary server Q_2 are set up, which form a computing cluster. When creating a shared section, you should know how it is distributed [12]. To generate a secret share, you only need to separate the numbers that need to be converted into two values. People Q_0 and Q_1 first exchange half of the shares, and then use the parts they hold to perform calculations and transactions, and finally get the final result. Q_0 will pass α_1 to Q_1, and Q_1 will pass β_0 to Q_0. Because Q_0 cannot enter β_1 in Q_1, it cannot determine the value of β. Adding is the most straightforward operation that can be performed through confidential sharing [13]. It can look like this:

$$\alpha + \beta = (\alpha_0 + \alpha_1) + (\beta_0 + \beta_1)$$

Equation (2.6) is rearranged by adding exchange rules and adding combination rules to:

$$\alpha + \beta = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1)$$

Q_0 solves for $\alpha_0 + \beta_0 \cdot Q_1$ solves $\alpha_1 + \beta_1$ to ensure that Q_0 gets only part of β. Q_1 gets only part of α. The two parties doing the multiplication need to communicate with each other during the operation [14]. The notation above is used to define multiplication using secret sharing in the following way:

$$\alpha\beta = (\alpha_0 + \alpha_1)(\beta_0 + \beta_1) = \alpha_0\beta_0 + \alpha_0\beta_1 + \alpha_1\beta_0 + \alpha_1\beta_1$$

Q_0 may be responsible for $\alpha_0\beta_0$ and Q_1 may be responsible for $\alpha_1\beta_1$. Giving intermediate project $(\alpha_0\beta_1 + \alpha_1\beta_0)$ to Q_0 or Q_1 will cause certain security risks. This is because either of the two items can be added together, revealing the original α or β. For example, if Q_0 wants to solve $\alpha_0\beta_1$, then Q_0 must have a β_1. Since Q_0 already contains β_0, we can find the value of β. Shielding can be used for concealment. When partial concealment is required, new unknowns can be introduced into the parties so that the mask can be eliminated without affecting the final result when all are calculated. The third party Q_2 generates information that it is unwilling to share with the other party for confidentiality purposes [15]. That is, masking $Q_0\beta_1$ and $Q_1\alpha_0$. In this article, the shields are called d and k, and ζ and η are called the values of the shield. Multiplying $\alpha\beta$ by Q_0 is:

$$f_0 = dk_0 + d_0\eta + \zeta k_0 + \zeta\eta$$

The multiplication of Q_1 by $\alpha\beta$ is equal to:

$$f_1 = dk_1 + d_1\eta + \zeta k_1$$

Builds the mask coefficient from $Q_2 \cdot Q_2$ generates 3 new numbers and divides them into different parts [16]. The first two digits are any number, and the third digit is the product of the first two digits. These values are obtained by subtracting the mask from the original data in the following way:

$$\zeta = (\alpha_0 - d_0) + (\alpha_1 - d_1)$$
$$\eta = (\beta_0 - k_0) + (\beta_1 - k_1)$$

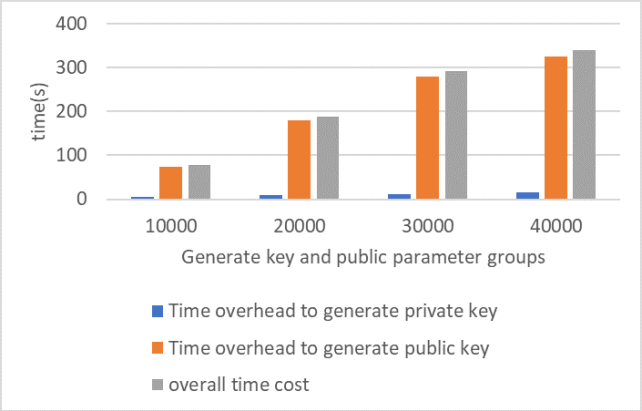

Fig. 3.1: *Key generation time cost.*

Q_2 transmits the values d_0 and k_0 to Q_0, and d_1 and k_1 to Q_1. Here the values of d_0 and k_0 take any integer, $d_1 = d - d_0, k_1 = k - k_0 \cdot Q_0$ then produces the $(\alpha_0 - d_0)$ part of ζ and the $(\beta_0 - k_0)$ part of $\eta \cdot Q_1$ produces $(\alpha_1 - d_1)$ regions of ζ and $(\beta_1 - k_1)$ regions of η; then D and G exchange the ζ and η shares without revealing any information about α or β; finally, the values are added to the corresponding equations. Combining (2.9) and (2.10) becomes:

$$f_0 + f_1 = \alpha\beta$$

Confidential sharing enables secure interaction with data. All expressions can be simulated using addition and homomorphic multiplication methods. Based on the two primitives of homomorphic addition and homomorphic multiplication, functions such as convolution, pooling, and approximation activation can be implemented, respectively. It can be applied to machine vision and linear or logistic regression fields. In practical applications, it is necessary to rewrite functions such as CNN, BP, MLP, logistic regression, etc., and use Taylor expansion approximation [17]. At the same time, the format of additive and multiplicative operations in homomorphic cryptography is maintained. This method is compatible with various mechanisms such as convolution, ReLU activation function, Maxpool and normalization, and can be combined with different inference methods to form corresponding safe computing solutions.

3. Analysis of experimental results.

3.1. The impact of Blowfish distance preservation on cryptography. The effect on the key is shown in Figure 3.1. The time it takes for a user to generate a key is mainly the time it takes to generate a key. The time when the private key was generated is not taken into account. The generation of a public-private key pair takes approximately eight milliseconds. Find the value of the transformation matrix during initialization keyword processing. Experiments have shown that only 30 fundamental transformations are needed. You can see from Figure 3.2 that encryption takes longer than decryption [18]. This is because generating random numbers that meet the requirements during cryptographic processing takes some time. The average time for each encoding is 0.4 milliseconds. Compared with existing algorithms, the algorithm proposed in this paper has better results. This is mainly because multiple bytes of data can be encrypted or decrypted simultaneously. Figure 3.3 shows that the calculation of 40,000 passwords only takes 67 seconds, and the calculation of 40,000 intermediate points takes only 67 seconds [19]. The amount of encryption required for clustering in a cloud computing environment is acceptable. Since cloud services have higher performance in practice, they can achieve higher privacy mining efficiency in practical applications.

3.2. The performance and accuracy of privacy-based cluster mining algorithms. The research content mainly involves the preprocessing time overhead of some users, part of the time overhead of cloud service providers, and problems compared with existing algorithms [20]. It can be seen from Figure 3.4 that

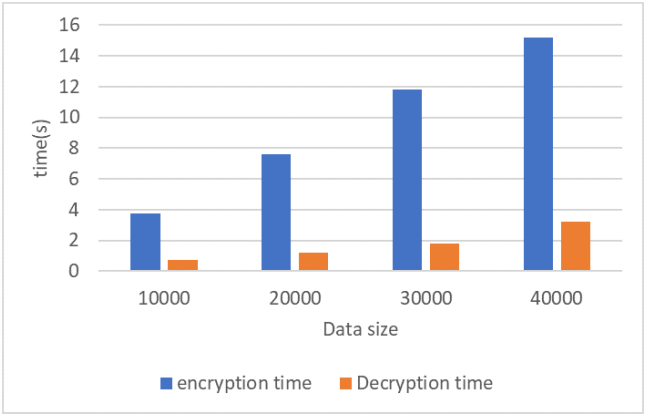

Fig. 3.2: *Encryption and decryption time.*

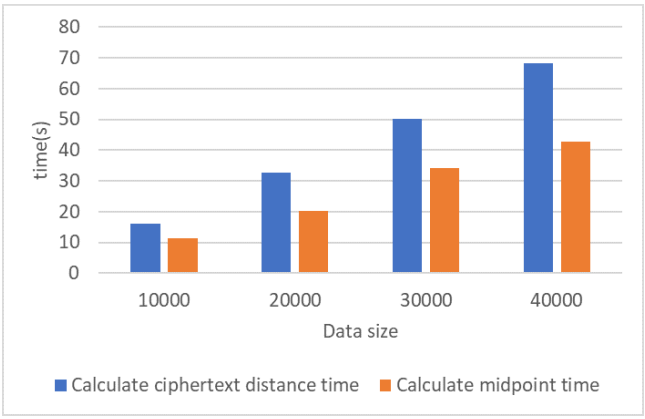

Fig. 3.3: *Computing time cost of ciphertext in the cloud.*

the data published by PPCM is more time-consuming than the existing algorithm, but it is still within the allowed range. The fundamental problem is that it is based on the Blowfish problem and increases the time cost required to keep the data confidential.

Figure 3.5 shows that the time complexity of the ciphertext k-means method is O(nlog(n)). The computational time complexity of the ciphertext hierarchical clustering algorithm is O(n2). The calculation time complexity of ciphertext DBSCAN is 0 (n). Research has found that in a natural environment, the calculation process of k-mean does not fully comply with the theoretical calculation complexity, and due to differences in sample type and initial value selection, The time consumption of the PPk-mean method may deviate from the analytical results to a certain extent. Experiments show that the privacy-based hierarchical mining method performs weakest, while the one based on PPDBscan is the best. This has a lot to do with the selected materials [21]. The other three methods also have advantages and disadvantages when processing various data types. In many situations, users need to perform multiple mining methods to discover potential clusters in a data set.

As shown in Figure 3.6, this method has the same accuracy as the current most correct RBT method but is safer. In exceptional circumstances, k unknown data sets may lead to inconsistent values of k. This project plans to use three methods to compare the k-means analysis results of the original data, and the PPk-means analysis results under the ciphertext to obtain the accuracy rate.

This algorithm can improve its security, accuracy, and mining efficiency. At the expense of this, the accuracy and security of users' private information are guaranteed. The accuracy of mining is very critical for users. If

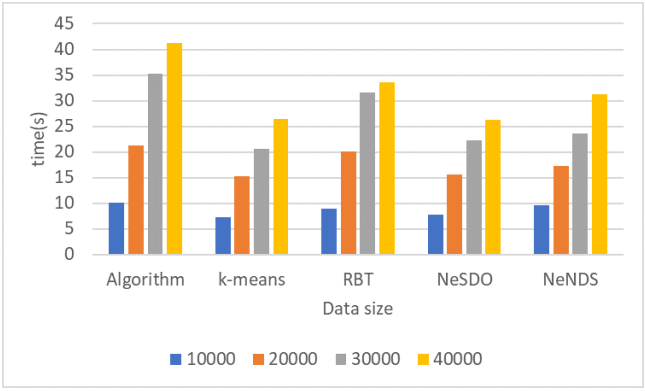

Fig. 3.4: *User preprocessing time overhead.*

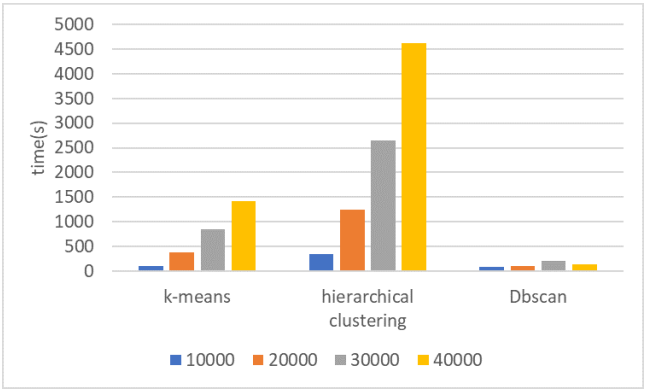

Fig. 3.5: . *Time cost of privacy-preserving cluster mining.*

there is a problem with the accuracy of discovery, it will likely bring irreparable consequences to the user.

4. Conclusion. The confidentiality of user information in ciphertext advertisements based on intelligent algorithms mainly focuses on the confidentiality of deduction data owned by users. This algorithm can keep the deduced data confidential throughout the entire reasoning process. Because encryption can only be performed by the data holder, confidentiality of the data and instant inference results can be considered simultaneously. And it also avoids the cost of repeated encryption and decryption. A model security method based on Blowfish is proposed. It can effectively store and read advertising mode documents while ensuring ciphertext security.

REFERENCES

[1] Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2021). Exploring motivations for online privacy protection behavior: Insights from panel data. Communication Research, 48(7), 953-977.
[2] Bleier, A., Goldfarb, A., & Tucker, C. (2020). Consumer privacy and the future of data-based innovation and marketing. International Journal of Research in Marketing, 37(3), 466-480.
[3] Shahabi, H. G., & Soni, S. (2023). SECURITY AND PRIVACY CHALLENGES IN VEHICULAR AD-HOC NETWORKS: THREATS, COUNTERMEASURES. Eigenpub Review of Science and Technology, 7(1), 22-38.
[4] Sharma, T., & Bashir, M. (2020). Use of apps in the COVID-19 response and the loss of privacy protection. Nature Medicine, 26(8), 1165-1167.
[5] Quan-Haase, A., & Ho, D. (2020). Online privacy concerns and privacy protection strategies among older adults in East York, Canada. Journal of the Association for Information Science and Technology, 71(9), 1089-1102.

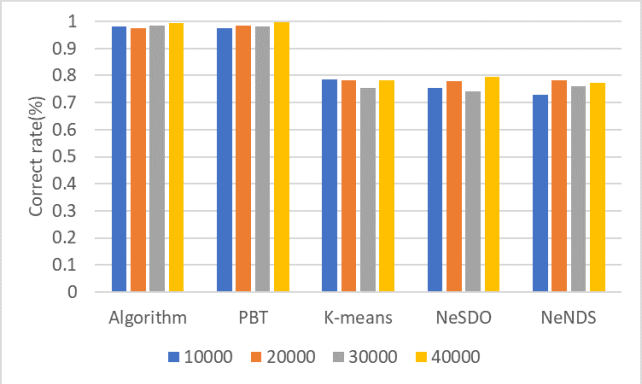

Fig. 3.6: *Privacy-preserving mining algorithm accuracy.*

[6] Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., & Lin, Z. (2021). When machine learning meets privacy: A survey and outlook. ACM Computing Surveys (CSUR), 54(2), 1-36.
[7] Jiang, H., Li, J., Zhao, P., Zeng, F., Xiao, Z., & Iyengar, A. (2021). Location privacy-preserving mechanisms in location-based services: A comprehensive survey. ACM Computing Surveys (CSUR), 54(1), 1-36.
[8] Zhou, X., Liang, W., She, J., Yan, Z., Kevin, I., & Wang, K. (2021). Two-layer federated learning with heterogeneous model aggregation for 6g supported internet of vehicles. IEEE Transactions on Vehicular Technology, 70(6), 5308-5317.
[9] Qu, Y., Pokhrel, S. R., Garg, S., Gao, L., & Xiang, Y. (2020). A blockchained federated learning framework for cognitive computing in industry 4.0 networks. IEEE Transactions on Industrial Informatics, 17(4), 2964-2973.
[10] Rafieian, O., & Yoganarasimhan, H. (2021). Targeting and privacy in mobile advertising. Marketing Science, 40(2), 193-218.
[11] Qiu, H., Qiu, M., Liu, M., & Memmi, G. (2020). Secure health data sharing for medical cyber-physical systems for the healthcare 4.0. IEEE journal of biomedical and health informatics, 24(9), 2499-2505.
[12] Duan, W., Gu, J., Wen, M., Zhang, G., Ji, Y., & Mumtaz, S. (2020). Emerging technologies for 5G-IoV networks: applications, trends and opportunities. IEEE Network, 34(5), 283-289.
[13] Du, M., Chen, Q., Xiao, J., Yang, H., & Ma, X. (2020). Supply chain finance innovation using blockchain. IEEE Transactions on Engineering Management, 67(4), 1045-1058.
[14] Siriwardhana, Y., Gür, G., Ylianttila, M., & Liyanage, M. (2021). The role of 5G for digital healthcare against COVID-19 pandemic: Opportunities and challenges. Ict Express, 7(2), 244-252.
[15] Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), 305-311.
[16] Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. American Economic Journal: Microeconomics, 14(4), 218-256.
[17] Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2022). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. IEEE journal of biomedical and health informatics, 27(2), 778-789.
[18] Qi, L., Hu, C., Zhang, X., Khosravi, M. R., Sharma, S., Pang, S., & Wang, T. (2020). Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment. IEEE Transactions on Industrial Informatics, 17(6), 4159-4167.
[19] Alazab, M., RM, S. P., Parimala, M., Maddikunta, P. K. R., Gadekallu, T. R., & Pham, Q. V. (2021). Federated learning for cybersecurity: Concepts, challenges, and future directions. IEEE Transactions on Industrial Informatics, 18(5), 3501-3509.
[20] Tan, T. M., & Saraniemi, S. (2023). Trust in blockchain-enabled exchanges: Future directions in blockchain marketing. Journal of the Academy of marketing Science, 51(4), 914-939.
[21] Stoilova, M., Livingstone, S., & Nandagiri, R. (2020). Digital by default: Children's capacity to understand and manage online data and privacy. Media and Communication, 8(4), 197-207.

Edited by: Zhigao Zheng
Special issue on: Graph Powered Big Aerospace Data Processing
Received: Dec 12, 2023
Accepted: Dec 29, 2023