Journal of King Saud University - Computer and Information Sciences 36 (2024) 102145

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University - Computer and Information Sciences

journal homepage: www.sciencedirect.com

Learning-driven Data Fabric Trends and Challenges for cloud-to-thing continuum

Praveen Kumar Donta [a,b,*], Chinmaya Kumar Dehury [c], Yu-Chen Hu [d]

[a] Distributed Systems Group, TU Wien, 1040 Vienna, Austria
[b] Department of Computer and Systems Sciences, Stockholm University, 16425 Kista, Stockholm, Sweden
[c] Mobile & Cloud Lab, University of Tartu, 50090 Tartu, Estonia
[d] Department of Computer Science, Tunghai University, Xitun District, Taichung City 407224, Taiwan, ROC

ARTICLE INFO

Keywords:
Cloud-to-thing continuum
Data fabric architecture
Machine learning and AI
Resource optimization

ABSTRACT

This special issue is a collection of emerging trends and challenges in applying learning-driven approaches to data fabric architectures within the cloud-to-thing continuum. As data generation and processing increasingly occur at the edge, there is a growing need for intelligent, adaptive data management solutions that seamlessly operate across distributed environments. In this special issue, we received research contributions from various groups around the world. We chose the eight most appropriate and novel contributions to include in this special issue. These eight contributions were further categorized into three themes: Data Handling approaches, resource optimization and management, and security and attacks. Additionally, this editorial suggests future research directions that will potentially lead to groundbreaking insights, which could pave the way for a new era of learning techniques in Data Fabric and the Cloud-to-Thing Continuum.

1. Introduction

The data fabric architecture can define the main structural component of our data analytics workflow, which we define as a homogeneous wrapper of data assets and services at the data layer. It can facilitate data representation, storing and processing, exchange and access, and data loss in a distributed manner (Sharma et al., 2023; Rieyan et al., 2024). It can be composed of a set of data fibers. Each data fiber is equipped with advanced data profiling and summarizing mechanisms and cloud-native capabilities at the resource layer, fostering rapid instantiation of data workflows over collected data where the data resides. It can also belong to one or more administrative domains (e.g., in multi-cloud setups) and must interconnect and interoperate (Dustdar et al., 2022). Moreover, they are highly dynamic and volatile, making it essential to manage their efficient and automated cloud-native orchestration at the infrastructure layer, including primitives such as dynamic provisioning or decommissioning, auto-scaling and migration, triggerable via declarative interfaces.

In this context, this special issue collate, and archive high-quality original research works from academic researchers and industry practitioners. Our aim is to explore the novel area of machine learning in various applications of Learning-driven Data Fabric Trends and Challenges for the Cloud-to-Thing Continuum. The primary technical research direction is to contribute to the Cloud-to-Edge/Things with little or no human intervention in data representation, storing and processing, exchange and access, data loss handling, and more. In addition, it aims to provide worldwide researchers and practitioners with an ideal platform to innovate new solutions targeting vital challenges. We received around 78 submissions within the deadline, around 63 papers were rejected by desk review and after review. Few papers were withdrawn and eight papers were accepted for publication. A summary of statistics of this special issue is provided in Fig. 1.

2. A thematic summary of the special issue

The accepted papers for this special issue are organized into three main themes: data handling, resource optimization, and security and

* Corresponding author at: Department of Computer and Systems Sciences, Stockholm University, 16425 Kista, Stockholm, Sweden.
E-mail addresses: praveen.donta@dsv.su.se (P.K. Donta), chinmaya.dehury@ut.ee (C.K. Dehury), ychu@thu.edu.tw (Y.-C. Hu).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2024.102145
Received 23 July 2024; Accepted 24 July 2024
Available online 30 July 2024

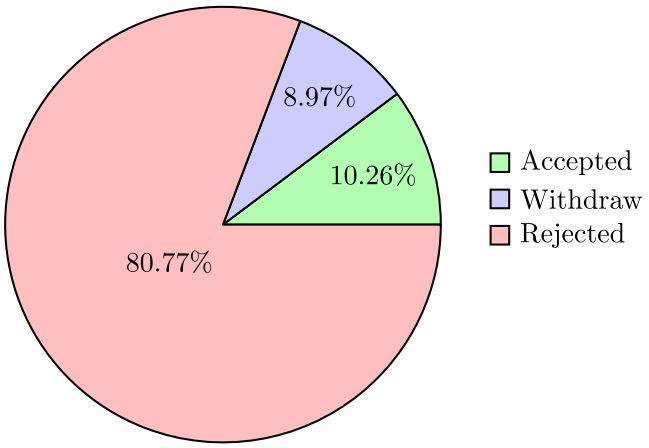

Fig. 1. Summary of Special issue submissions and acceptance.

attacks. We received three contributions focusing on various data handling approaches. These papers discuss innovative methods and techniques for managing and processing data efficiently. Two articles were published on resource optimization and management. These works explore strategies for optimizing resource use, ensuring efficient and effective management. Three more papers address handling attacks and providing secure mechanisms. These contributions cover various aspects of security, including identifying and mitigating attacks, and implementing robust security measures.

2.1. Data handling (Theme 1)

In Yi (2023), *H. Yi* implemented an efficient learning-driven data transmission algorithm for cloud-to-thing continuum. Using Forward Error Correction (FEC) and Gaussian Elimination (GE), *Yi* recovered lost data packets at the application layer and optimized transmission efficiency within the cloud-to-edge continuum. This work also ensures continuous learning along with an online decision-making capabilities. In this way, cloud-based AI systems and edge devices can more efficiently communicate and manage Internet traffic loads.

Data extraction from larger scenes is more challenging than smaller ones. To address this, Li et al. (2024) introduce an innovative approach designed to efficiently obtain high-quality scene data and accurately model targets. The contributions in this paper include an adaptive density-based voxel grid filtering algorithm and a probability statistics histogram method implemented during preprocessing to enhance data quality. Additionally, they propose a novel method for automatic extraction of target PC data through an Adjacent Feature Plane Constraint (AFPC) clustering technique. Initially, this approach uses experiential and statistical height attributes to determine the height of the target object's point cloud. Then, cluster points corresponding to each height using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, which clusters points based on their density distribution. These contributions collectively improve the efficiency and accuracy of extracting target PC data from larger and more complex scenes.

Shang et al. (2024) introduced a rule-based multi-class classification method for multi-source tabular data in cloud–edge systems, outperforming seven leading explainable methodologies over six public datasets. This method combines strong predictive performance with clear insights, achieving these results without preprocessing, even for complex datasets. The Wilcoxon rank test performs better when missing values are present in comparison with other methods. Through experiments, this method confirmed its effectiveness through convergence analysis and iterative visualization. The source code for this approach is available on GitHub[1] for the purpose of reproducibility, reusability, extensibility of their research result. Thanks to *J. Shang et al.*

2.2. Resource optimization (Theme 2)

Nauman et al. (2024) contributed a paper entitled "Communication and computational resource optimization for Industry 5.0 smart devices empowered by MEC". This study presents an integrated framework for Mobile Edge Computing (MEC) and Industry 5.0, enhancing industrial productivity, effectiveness, and adaptability. This joint optimization problem improves computational energy efficiency by optimally distributing resources like processing power. This work also addressed the complex problem by breaking it down into smaller, solvable parts and performing the computations in a distributed manner. This hybrid approach, incorporating partial offloading, outperforms other methods, highlighting its suitability for joint optimization. This pioneering study demonstrates MEC's potential to revolutionize Industry 5.0, enabling real-time data processing, reduced communication latency, faster response times, and increased system reliability, addressing the limitations of smart devices with short battery lives and limited computational power.

Zhang et al. (2024) contributed an article entitled "Resource allocation strategy based on service function Chaining in Multi-Access edge computing network" for our special issue. This article presents a MEC network system model focusing on terminal device task processing and service provisioning. Through the use of Network Function Virtualization (NFV) technology, two strategies for optimizing Service Function Chaining (SFC) resources are proposed by transforming the resource allocation problem into a Virtualized Network Function (VNF) node deployment cost optimization problem. A deep reinforcement learning-based model is introduced for dynamic SFC, optimizing resource allocation through value and behavior policies. Model evaluation and network simulation experiments demonstrate that the proposed method improves the SFC task processing success rate by over 30% and reduces resource allocation costs by 20% compared to other methods. *X. Zhang et al.* confirmed through experiments in various scenarios that this model is highly effective for general MEC network task processing.

2.3. Security and attacks (Theme 3)

Ye and Yang (2024) contributed a paper entitled "An ECC with error detection and against side channel attacks for resource constrained devices". This work introduces an improved elliptic curve encryption scheme tailored for resource-constrained IoT devices, featuring an error detection and correction mechanism. The enhanced Montgomery ladder algorithm facilitates error detection, while re-computation handles error correction. To protect against side-channel attacks, base point blinding via randomization is employed. This scheme achieves a 100% error detection rate and robustly defends against SSCA, DPA, RPA, ZPA, and Relative Doubling Attack. This scheme is particularly suitable for IoT devices with limited resources, addressing electronic interference and communication attacks with improved efficiency and security. The source code for this approach is available on GitHub[2] for the purpose of reproducibility, reusability, extensibility of their research result. Thanks to *J. Ye* and *Z. Yang*.

Wang et al. (2024) contributed to research entitled "Blockchain user digital identity big data and information security process protection based on network trust" in this special issue. This research pioneers cross-disciplinary integration in digital identity and network trust through the NOPI (Network trust, Organizational collaboration,

[1] https://github.com/wenge963/MCC_cloudbased
[2] https://github.com/yangzewen/MLEDC

Personal big data, Information technology) model, bridging social psychology, data management, and information security. It explores the relationships between network trust, blockchain organization, user identity, big data, and information security, proposing precise digital identity attributes and management chains to enhance blockchain security. Introducing dynamic security mechanisms and collaborative governance processes, it offers effective digital identity governance solutions. According to *F. Wang et al.*, this research facilitates reliable digital identity integration into financial services, improves employment opportunities, and aids governments and enterprises in fraud prevention and service quality. Further, collaborative identity management enhances cross-border transactions' efficiency and international cooperation, fostering global trust.

Alshahrani et al. (2024) contributed to advancing ubiquitous learning through improved blockchain integration and incentive strategies in the paper entitled "Enabling intrinsic intelligence, ubiquitous learning and blockchain empowerment for trust and reliability in 6G network evolution" to this special issue. This study explores the integration of 6G wireless communication networks and blockchain technology to enhance ubiquitous learning in the cloud–edge paradigm. This work addresses various challenges including varied communication settings and resource limitations on blockchain nodes. This work considers a contract theory-based incentive mechanism and a Byzantine fault tolerance mechanism tailored to ubiquitous learning. The model optimizes block rewards for both symmetric and asymmetric information scenarios, enhancing reliable intelligence and performance in 6G networks.

We believe that these **EIGHT** papers make significant contributions and offer valuable insights into the challenges and opportunities of "Learning-driven Data Fabric Trends and Challenges for Cloud-to-Thing Continuum". As a result of this special issue, we hope that more research will be conducted in this area to address Data Fabric Trends and Challenge with respect to the Cloud-to-Thing Continuum paradigm and applying novel learning techniques.

3. Future research directions

Our special issue covers only a limited number of issues, whereas there is huge scope for further research in this field. We list the most emerging open research issues to be addressed in the fields of data fabric and cloud-to-things continuum, and applicability of learning-driven approaches.

- **Scalable Learning**: Need for develop scalable and efficient learning strategies that are optimized for distributed environments, enabling robust performance and real-time processing capabilities across diverse and resource-constrained computing continuum (Pujol et al., 2024). For instance, a deep neural network implemented on a GPU-based server may outperform the same model implemented on a CPU-based server due to higher computing power and parallel processing (Li and Wang, 2024).
- **Adaptability at scale**: Need for adaptive data fabric architectures that dynamically adjust to changing network conditions, data volumes, and processing requirements, and ensuring optimal performance through efficient resource utilization. For example, in high-traffic periods where data volumes are exceptionally high, an adaptive data fabric architecture can allocate additional resources to handle the increased load and ensure smooth data processing. Similarly, in situations where processing requirements are more complex, such as running advanced analytics (for example, deep learning algorithms), an adaptive architecture can dynamically allocate the necessary computational power to achieve optimal performance (Macías et al., 2024).
- **Context awareness**: Investigate context-aware data management strategies that consider contextual information of devices and applications, improving data-driven decisions' relevance and accuracy. Implementing context-aware data management strategies can be challenging due to several factors (Chhetri et al., 2024). One challenge is the need for a comprehensive and accurate understanding of the context in which the data is collected. This requires the integration of various data sources and the development of sophisticated algorithms to analyze and interpret contextual information (Fadlallah et al., 2023; Vila et al., 2023). For instance, context-aware data management strategies require an understanding of user behavior, the device used to access the data, and the environment in which the data is being collected (Ochoa et al., 2023).
- **Real-time data analytics**: Providing actionable insights and prompt decision-making to cloud-to-thing data streams through the development of real-time data analytics solutions. Real-time data analytics solutions work by continuously monitoring and analyzing data as it is generated from cloud-to-thing data streams (Sedlak et al., 2023). This allows for immediate detection of patterns, trends, and anomalies, providing organizations with actionable insights and enabling prompt decision-making based on up-to-date information.
- **Data Integration**: Enhance data consistency, quality, and interoperability across the cloud-to-thing continuum by investigating advanced methods for seamless integration of heterogeneous data sources. By exploring more advanced integration methods organizations can ensure that data is consistent, of high quality, and interoperable across the cloud-to-thing continuum (Zhu et al., 2023; Paneque et al., 2023). This will help them better understand their data and make better decisions (Thakkar et al., 2023).
- **Role of Large Language Models**: Large language models (LLMs) are advanced AI systems that utilize vast amounts of text data to understand, generate, and respond to human language with high levels of fluency and contextual accuracy. The future research should also explore the potential of deploying compact and efficient LLMs specifically designed for resource-constrained devices within the cloud-to-thing continuum. These models can enhance data processing, analysis, and decision-making by performing natural language processing tasks, automated summarization, and contextual understanding. The focus should be on developing lightweight versions of LLMs that maintain high performance while being small enough to operate on edge devices with limited computational resources. This would enable more intuitive human–computer interactions, improve the efficiency of data-driven insights, and enable real-time recommendations and report generation, optimizing resource use and enhancing overall system intelligence.

CRediT authorship contribution statement

Praveen Kumar Donta: Conceptualization, Data curation, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft. **Chinmaya Kumar Dehury:** Formal analysis, Investigation, Methodology, Resources, Writing – review & editing. **Yu-Chen Hu:** Data curation, Writing – review & editing.

Declaration of competing 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 paper.

Acknowledgments

We are extremely grateful to academic experts who voluntarily gave their time and contributed works to maintaining the high quality standards of the Journal. We would like to sincerely thank all reviewers who participated in the manuscript assessment process and timely

returning the review reports to made decisions timely for this special issue. We sincerely thanks to *Prof. Nasser-Eddine Rikli* (Editor-in-Chief) for allowing us to organize this special issue, and supporting throughout the special issue. We also thank the editorial office of the Journal of King Saud University Computer and Information Sciences, Elsevier Journal who were instrumental in providing a rapid and efficient editorial process. Special thanks to *Caroline Zhang* for timely response and support throughout the special issue.

References

Alshahrani, R., Shabaz, M., Khan, M.A., Kadry, S., 2024. Enabling intrinsic intelligence, ubiquitous learning and blockchain empowerment for trust and reliability in 6G network evolution. J. King Saud Univ.-Computer Inf. Sci. 36 (4), 102041.

Chhetri, T.R., Dehury, C.K., Varghese, B., Fensel, A., Srirama, S.N., DeLong, R.J., 2024. Enabling privacy-aware interoperable and quality IoT data sharing with context. Future Gener. Comput. Syst. 157, 164–179. http://dx.doi.org/10.1016/j.future.2024.03.039, URL https://www.sciencedirect.com/science/article/pii/S0167739X24001109.

Dustdar, S., Pujol, V.C., Donta, P.K., 2022. On distributed computing continuum systems. IEEE Trans. Knowl. Data Eng. 35 (4), 4092–4105.

Fadlallah, H., Kilany, R., Dhayne, H., El Haddad, R., Haque, R., Taher, Y., Jaber, A., 2023. Context-aware big data quality assessment: a scoping review. ACM J. Data Inf. Qual. 15 (3), 1–33.

Li, R., Gan, L., Di, Y., Tian, H., Zuo, Q., Luo, Y., Wu, X., Wang, H., 2024. Neighborhood constraint extraction for rapid modeling of point cloud scenes in large-scale power grid substations. J. King Saud Univ.-Computer Inf. Sci. 101963.

Li, J., Wang, D., 2024. Stochastic configuration networks with CPU-GPU implementation for large-scale data analytics. Inform. Sci. 667, 120497.

Macías, A., Muñoz, D., Navarro, E., González, P., 2024. Data fabric and digital twins: An integrated approach for data fusion design and evaluation of pervasive systems. Inf. Fusion 103, 102139.

Nauman, A., Khan, W.U., Aldehim, G., Alqahtani, H., Alruwais, N., Al Duhayyim, M., Dev, K., Min, H., Nkenyereye, L., 2024. Communication and computational resource optimization for industry 5.0 smart devices empowered by MEC. J. King Saud Univ.-Computer Inf. Sci. 36 (1), 101870.

Ochoa, W., Larrinaga, F., Pérez, A., 2023. Context-aware workflow management for smart manufacturing: A literature review of semantic web-based approaches. Future Gener. Comput. Syst. 145, 38–55.

Paneque, M., del Mar Roldán-García, M., García-Nieto, J., 2023. e-LION: Data integration semantic model to enhance predictive analytics in e-Learning. Expert Syst. Appl. 213, 118892.

Pujol, V.C., Sedlak, B., Donta, P.K., Dustdar, S., 2024. On causality in distributed continuum systems. IEEE Internet Comput. 28 (2), 57–64.

Rieyan, S.A., News, M.R.K., Rahman, A.M., Khan, S.A., Zaarif, S.T.J., Alam, M.G.R., Hassan, M.M., Ianni, M., Fortino, G., 2024. An advanced data fabric architecture leveraging homomorphic encryption and federated learning. Inf. Fusion 102, 102004.

Sedlak, B., Murturi, I., Donta, P.K., Dustdar, S., 2023. A privacy enforcing framework for data streams on the edge. IEEE Trans. Emerg. Top. Comput. http://dx.doi.org/10.1109/TETC.2023.3315131.

Shang, J., Xiao, Z., Tao, T., Wang, J., Wu, Z., 2024. A heuristic method for discovering multi-class classification rules from multi-source data in cloud–edge system. J. King Saud Univ.-Computer Inf. Sci. 101962.

Sharma, V., Balusamy, B., Thomas, J.J., Atlas, L.G., 2023. Data Fabric Architectures: Web-Driven Applications. Walter de Gruyter GmbH & Co KG.

Thakkar, H.K., Dehury, C.K., Sahoo, P.K., Veeravalli, B. (Eds.), 2023. Predictive Analytics in Cloud, Fog, and Edge Computing: Perspectives and Practices of Blockchain, IoT, and 5G. Springer, Cham, http://dx.doi.org/10.1007/978-3-031-18034-7.

Vila, M., Sancho, M.-R., Teniente, E., Vilajosana, X., 2023. Critical infrastructure awareness based on IoT context data. Internet of Things 23, 100855.

Wang, F., Gai, Y., Zhang, H., 2024. Blockchain user digital identity big data and information security process protection based on network trust. J. King Saud Univ.-Computer Inf. Sci. 102031.

Ye, J., Yang, Z., 2024. An ECC with error detection and against side channel attacks for resource constrained devices. J. King Saud Univ.-Computer Inf. Sci. 102019.

Yi, H., 2023. Efficient learning-driven data transmission algorithm for cloud-to-thing continuum. J. King Saud Univ.-Computer Inf. Sci. 35 (10), 101834.

Zhang, X., Huang, Z., Huang, L., Yang, H., 2024. Resource allocation strategy based on service function chaining in multi-access edge computing network. J. King Saud Univ.-Computer Inf. Sci. 102001.

Zhu, X., Wang, L., Zhou, C., Cao, X., Gong, Y., Chen, L., 2023. A survey on deep learning approaches for data integration in autonomous driving system. arXiv preprint arXiv:2306.11740.

Praveen Kumar Donta (Senior Member, IEEE'22), currently Senior Lecturer at the Department of Computer and Systems Sciences, Stockholm University, Sweden. He worked at the Distributed Systems Group at TU Wien, as a Postdoctoral researcher from July 2021 to June 2024. He received his Ph.D. from the Indian Institute of Technology (Indian School of Mines), Dhanbad, in the Department of Computer Science & Engineering in June 2021. He was a Visiting Ph.D. fellow at Mobile&Cloud Lab, University of Tartu, Estonia from July 2019 to Jan 2020. He received his Master in Technology and Bachelor in Technology from the Department of Computer Science and Engineering at JNTUA, Ananthapur, with Distinction in 2014 and 2012, respectively. He is an IEEE Senior Member and an ACM Professional Member. He is serving as Editorial board member in Computing Springer, ETT Wiley, POLS One, Measurement, and Computer Communications Elsevier Journals. His current research is on learning-driven distributed computing continuum systems, Casual and Conscious Continuum Systems, and Intelligent Data Protocols.

Chinmaya Kumar Dehury received bachelor degree from Sambalpur University, India, in June 2009 and MCA degree from Biju Pattnaik University of Technology, India, in June 2013. He received the Ph.D. Degree in the department of Computer Science and Information Engineering, Chang Gung University, Taiwan. Currently, he is a postdoctoral research fellow in the Mobile & Cloud Lab, Institute of Computer Science, University of Tartu, Estonia. His research interests include scheduling, resource management and fault tolerance problems of Cloud and fog Computing, and the application of artificial intelligence in cloud management. He is an reviewer to several journals and conferences, such as IEEE TPDS, IEEE JSAC, Wiley Software: Practice and Experience, etc.

Yu-Chen Hu (Senior Member, IEEE'15) received his PhD. degree in computer science and information engineering from the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan in 1999. Currently, he is a professor at the Department of Computer Science, Tunghai University, Taiwan. Dr. Hu worked as a professor in the Department of Computer Science and Information Management, Providence University, Taichung, Taiwan. He is a senior member of IEEE. He is also a member of Phi Tau Phi Society of the Republic of China. His research interests include image and signal processing, data compression, information security, data engineering, and computer networks.