IntechOpen

Edge Computing

Architecture and Applications for Smart Cities

Edited by Yu Chen and Ronghua Xu

Edge Computing - Architecture and Applications for Smart Cities

Edited by Yu Chen and Ronghua Xu

Published in London, United Kingdom

Edge Computing – Architecture and Applications for Smart Cities
http://dx.doi.org/10.5772/intechopen.1002562
Edited by Yu Chen and Ronghua Xu

Contributors
Alexios Birbas, Alex Papalexopoulos, Anna Tzanakaki, Auhood Al-Hossenat, Balaji Venkatesalu Ramasamy, Christina (Tanya) Politi, Christos Tranoris, Deeraj Nagothu, Dinesh Kumar Jayaraman Rajendiran, Eleftherios Mylonas, Ioannis Moraitis, Jesús Gutiérrez Terán, Joby Titus T, Karthi Samiyampalayam Palanisamy, Mouiad Al-Wahah, Nikolaos Tzanis, Panagiotis Papaioannou, Ronghua Xu, Sheng Chen, Spyros Denazis, Stella Bvuma, Visvesvaran Chandramohan, Xiaoyi Tao, Xin Xie, Xiulong Liu, Yu Chen, Yu Chen

First published in London, United Kingdom, 2024 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 167-169 Great Portland Street, London, W1W 5PF, United Kingdom

For EU product safety concerns: IN TECH d.o.o., Prolaz Marije Krucifikse Kozulić 3, 51000 Rijeka, Croatia, info@intechopen.com or visit our website at intechopen.com.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Edge Computing - Architecture and Applications for Smart Cities
Edited by Yu Chen and Ronghua Xu
p. cm.
Print ISBN 978-0-85466-770-3
Online ISBN 978-0-85466-769-7
eBook (PDF) ISBN 978-0-85466-771-0

If disposing of this product, please recycle the paper responsibly.

Meet the editors

Dr. Yu Chen is an Electrical and Computer Engineering Professor at Binghamton University in Binghamton, USA. He received a Ph.D. in Electrical Engineering from the University of Southern California (USC) in 2006. His research centers on Internet of Things (IoT) technologies and their applications in creating intelligent and interconnected environments. Dr. Chen's publications include over 200 papers published in scholarly journals, conference proceedings, and books. NSF(National Science Foundation), DoD(Department of Defense), AFOSR(Air Force Office of Scientific Research), AFRL(Air Force Research Laboratory), New York State, and industrial partners have funded his research. He has served as a reviewer for NSF panels, the DoE(Department of Energy) Independent Review Panel, international journals, and the Technical Program Committee (TPC) of prestigious conferences. Dr. Chen is a Fellow of SPIE(Society of Photo-Optical Instrumentation Engineers), a Senior Member of ACM(Association for Computing Machinery) and IEEE(Institute of Electrical and Electronics Engineers), and a SIGMA XI member.

Dr. Ronghua Xu is an Assistant Professor of Applied Computing at Michigan Technological University in the USA. He received a Ph.D. in Electrical Engineering from Binghamton University in 2023. His research focuses on constructing intelligent and resilient networked systems through the organic synergy of the Internet of Things (IoT), Machine Learning (ML), and Blockchain technologies. Dr. Xu's publications include over 50 papers published in scholarly journals, conference proceedings, and books. He has served as a reviewer for international journals and the Technical Program Committee (TPC) of prestigious conferences. Dr. Xu is a Member of ACM(Association for Computing Machinery) and IEEE(Institute of Electrical and Electronics Engineers).

Contents

Preface

We live in an age when rapid technological advancement has brought edge computing to the foreground of digital evolution. Now more than ever, there is an increasing need for the ability to process data in real-time with secure ways and robust computational power as cities get smarter and industries grow interconnected. This book, *Edge Computing – Architecture and Applications for Smart Cities*, explores the cutting-edge advancements and practical applications of edge computing in today's dynamic technological landscape.

Starting with an introductory chapter, the book is organized into two main sections: Novel Architectures and Application Scenarios. Each section comprises meticulously researched chapters addressing critical edge computing aspects.

Section I: Novel Architectures

Chapter 1: Introductory Chapter: Edge Computing in the Evolution of Smart Cities

Authored by Yu Chen and Ronghua Xu, this introductory chapter provides a high-level, concise overview of edge computing among smart cities, including its history, current status, and future perspectives. This chapter serves readers as an entry point into a vast, ever-evolving space of edge computing in smart cities.

Chapter 2: AR-Edge: Autonomous and Resilient Edge Computing Architecture for Smart Cities

Authored by Ronghua Xu, Deeraj Nagothu, and Yu Chen, this chapter introduces an Autonomous and Resilient Edge (AR-Edge) computing architecture that integrates AI, software-defined network (SDN), and Blockchain technologies to enable next-generation edge computing networks. The AR-Edge aims to provide autonomous, secure, and resilient edge networks for dynamic and complex IoT ecosystems. A preliminary proof-of-concept prototype of an intelligent transportation system (ITS) demonstrates the feasibility of applying AR-Edge in real-world scenarios.

Chapter 3: An Effective and Efficient Computation Architecture for Edge Computing Devices on IoMT-Based Deep Belief Networks

The authors, Dinesh Kumar Jayaraman Rajendiran et al., elaborate on the Internet of Medical Things (IoMT). They proposed a sustainable approximation adder architecture to improve the performance of edge devices in IoMT applications, focusing on faster computation and reduced latency in critical scenarios such as robotic surgery and autonomous vehicles.

Section II: Application Scenarios

Chapter 4: Safety Assurance in IoT-Based Smart Homes

Mouiad Al Wahah and Auhood Al-Hossenat analyze the safety technologies essential for smart home systems in this chapter. They review current trends in security-enabled

safety monitoring frameworks for IoT-based smart homes, emphasizing the importance of operational data analysis and real-time safety assurance.

Chapter 5: The Impact of 5G-Enabled Edge-Cloud Services on Energy Facilities in Industry 4.0

This chapter, authored by Nikolaos Tzanis et al., investigates the transformative impact of 5G and edge-cloud services on industrial energy facilities. It explores how these technologies drive efficiency and innovation in Industry 4.0 settings.

Chapter 6: Service Provision and Price Strategies in Edge Computing

Authored by Xiulong Liu, Xiaoyi Tao, Sheng Chen, and Xin Xie, this chapter examines strategic deployment models for edge computing services. It discusses dynamic pricing frameworks and service placement mechanisms that optimize social welfare and system performance while balancing economic objectives and quality of service.

Chapter 7: Understanding Citizen Engagement in the Era of Smart Cities

In this chapter, Stella Bvuma explores the importance of public participation in developing smart cities. This chapter emphasizes tackling disenfranchisement, fostering digital literacy, and utilizing diverse engagement strategies to ensure inclusive, cooperative, and equitable smart city development.

Each chapter in this book is penned by experts who bring invaluable insights and practical knowledge to the forefront of edge computing. We aim to provide a comprehensive overview of recent developments, challenges, and future research directions in edge computing to help readers get inspired and navigate various possibilities regarding this disruptive technology.

We hope this book will provide the basis for researchers, practitioners, and policymakers working toward realizing the full promise of edge computing. As we move deeper into the 21st century, the frontiers of technology appear limitless, and with them, so do opportunities for improving our cities through innovation.

Thanks for being on this journey with us.

Yu Chen
Department of Electrical and Computer Engineering,
Binghamton University,
Binghamton, New York, USA

Ronghua Xu
Department of Applied Computing,
Michigan Technological University,
Houghton, Michigan, USA

Section 1

Novel Architectures

Chapter 1

Introductory Chapter: Edge Computing in the Evolution of Smart Cities

Yu Chen and Ronghua Xu

1. Introduction

While the precise moment when the concept of "smart cities" was officially introduced is still debatable, we have witnessed a remarkable transformation in past decades [1]. Today, based on the convergence of advanced technologies, Smart cities have become a reality, enfolding urban landscapes across almost all parts of the globe. One of the primary drivers is edge computing [2]. The smart city model instinctively embeds information and communication technologies in the heart of urban existence. Smart urban ecosystems boost residents' quality of life, help optimize resource consumption, and provide better solutions in terms of sustainability and resilience [3, 4]. The issues across cities nowadays are becoming more and more complex as the urban population grows, such as traffic congestion [5], environmental concerns [6], and public safety [7]; hence, there is a growing need of more advanced technological solutions.

Edge computing, a distributed computing paradigm, brings data processing and storage closer to where the data are generated. This approach allows a significant migration of computation capability from the remote centralized cloud centers to the network edge. By processing data at or near its source, edge computing addresses several critical limitations of cloud-based systems, particularly in the context of smart cities [8]. Processing data in real-time at the edge allows faster responses, less network congestion, and more robust privacy protection. These are among the key considerations for fast-moving cities that rely on large amounts of data [9]. For instance, edge-enabled intelligent traffic management systems can adapt in real-time to changing road conditions [10], while smart grids ensure the most efficient localized energy distribution [11].

The widespread deployment of intelligent Internet of Things (IoT) devices in smart cities further contributes to the overwhelming data influx. Streetlights, parking meters, and various environmental sensors generate enormous amounts of data [12]. A computational framework is required to handle this data tsunami, transforming raw input information into actionable insights at the urban edge. With fifth-generation (5G) communication technology becoming mainstream, the scope of edge computing in smart cities expands. 5G networks align with the requirements of edge computing, providing high-speed, low-latency capabilities that enable more intelligent and control-driven urban systems [13]. This symbiosis of technologies opens up new opportunities, such as autonomous vehicles, augmented reality-supported public services, and hyperlocal environmental sensing [14].

It is widely recognized that the evolution of smart cities, powered by edge computing, represents an entirely new vision of how we design, build, and govern our

urban environments [15]. In the future, the intelligent edge will give rise to smart, adaptive, and resilient cities with responsive inhabitants.

2. Current state of edge computing in urban environments

As edge computing matures, its implementation in urban environments is rapidly evolving. This section provides a brief overview of the current landscape of edge computing in smart cities, highlighting some existing implementations, success stories, challenges, and some lessons.

Edge computing is already effectively deployed in numerous smart city applications to improve quality of life in various ways. For instance, cities like Singapore, San Francisco, and Barcelona utilize edge computing solutions to manage traffic in real-time through optimized traffic flow strategies, leading to reduced congestion [16]. Chicago employs edge-based video analytics for public safety, enabling real-time threat assessments and quicker response times [17]. New York City's LinkNYC project replaced phone booths with Wi-Fi kiosks incorporating edge computing capabilities, providing citizens with real-time information and services [18]. Amsterdam uses edge computing for air quality monitoring and noise pollution control, allowing for rapid responses to environmental changes [19]. A purpose-built smart city in Songdo, South Korea, extensively utilizes edge computing in its urban management systems, demonstrating the potential of integrated edge solutions in greenfield urban developments [20]. Additionally, cities like Oslo implement edge computing in power distribution systems, resulting in smarter grids and improved fault detection [21].

While we observed many successful use cases and the potential of edge computing in urban settings, there are still challenges that hinder its widespread adoption:

- *Standardization*: The lack of unified standards for edge computing architectures and protocols hinders interoperability and scalability across different smart city systems [22].

- *Security concerns*: The distributed nature of edge computing introduces new security vulnerabilities that must be addressed to protect sensitive urban data [9].

- *Energy efficiency*: Powering numerous edge devices and ensuring their efficient operation remains a significant challenge, particularly in resource-constrained urban environments [21].

- *Data governance*: Issues surrounding data ownership, privacy, and regulatory compliance are becoming increasingly complex as more data is processed at the edge [9, 23].

- *Infrastructure integration*: Retrofitting existing urban infrastructure to accommodate edge computing capabilities can be costly and logistically challenging [22].

The extensive examples of smart city projects that have been documented now, whether they are termed successful or not, provide aspects for us to learn from in future implementations. They underscore the need to involve stakeholders and citizens in the co-design and deployment of edge computing solutions [24]. They also

point to the necessity for flexible and scalable architectures to evolve along with rapidly evolving urban needs and technological developments. Ensuring public-private partnerships is another key to further driving innovation and overcoming implementation bottlenecks. As the landscape for edge computing continues metamorphosizing over time, in relation to urban areas, these initiatives serve as a reminder of how promising this transformation could be and simultaneously expose us to opportunities that innovation and concerted efforts are critical to unleashing its full potential.

3. The future of intelligent edge computing in smart cities

As we look to the future, the convergence of edge computing with advanced artificial intelligence (AI) and machine learning (ML) technologies promises to usher in a new era of intelligent urban systems. This section explores the emerging trends and potential future developments in intelligent edge computing for smart cities, highlighting the transformative potential of these technologies.

Integrating edge computing and the advancing AI and ML algorithms will significantly enhance smart city systems [25, 26]. This will also further improve the ability to handle real-time data for processing and decision-making in future edge systems, which, through the use of AI, would be able to process immense amounts of urban data instantaneously so that critical city functions can respond in real-time. Thanks to predictive analytics enabled by AI, as another example, urban planning will become more effective in addressing the needs of a city and its residents by allowing for forward-thinking strategies that anticipate transportation trends, energy usage, or public health concerns. In addition, adaptive learning systems will equip intelligent edge systems to keep learning from and adjusting to the shifting urban landscape in real-time, enabling continual optimization of city operations with no human touch required.

The future of smart cities is autonomous, self-organizing edge computing systems [27, 28]. Based on self-healing network infrastructures, faults will be detected autonomously, diagnosed, and repaired automatically through the restoration process for robust and resilient urban services. This implies decentralized decision-making systems will spread the intelligence out over the city, decreasing dependence on central authorities and leading to solutions that better meet the needs of individual neighborhoods. However, to see things take off, we need to ensure complete autonomy in not just a connected pool of self-driving cars but also urban service optimization where edge-based AI manages systems like traffic and waste management at the city level, adapting continuously based on real-time aspects for seamless city operations.

One significant development on the horizon is cognitive edge computing that allows for the human aspect of understanding and reasoning by the systems [29, 30]. The mapmaking process hinges on many data sources, such as satellite imagery, zip code layers, local demographics, and traffic patterns. These context-aware computing systems make sense of the sprawling urban landscape using variables like cultural events (which influence congestion), current weather (which influences pizza delivery), and who is out tonight with their friends from work (so only show bar options to Joe's that involve hockey). They will also add a semantical understanding of city dynamics: a more profound, nuanced comprehension of what happens in the city. Consequently, such a system will be able to address complex urban problems with ethical considerations and long-term likely effects in a human-like way of reasoning.

Additionally, several cutting-edge technologies would enhance intelligent edge computing capabilities even more. Quantum computing at the edge will dramatically expand compute resources to handle complex calculations essential for urban optimization [31]. Similarly, neuromorphic computing, based on the brain, which has adaptive capability and resistance, is leading to edge systems that resemble nature's ecosystem [32]. Blockchain and distributed ledger technologies could further improve security, transparency, and decentralization in the context of resource sharing or citizen services [33, 34]. The technologies will increase the capabilities of edge systems to deliver more secure and effective urban services.

This convergence of maturing technologies points to smart cities that are becoming better not just at responding efficiently but also at being truly intelligent and adaptive. At the intelligent edge, cities can begin to act as the cognitive backbone of smart city environments, acting as a thinking, learning, decentralized system process previously only imagined or through science fiction.

Still, it will not come without struggle. This requires sharpening ethical standards, addressing genuine worries about privacy, and introducing more resilient governance frameworks [35]. Intelligent edge computing systems that benefit urban residents require interdisciplinary efforts between technologists, urban planners, policymakers, and citizens.

4. Societal implications of intelligent edge computing

As intelligent edge computing continues reshaping smart cities, its impact extends beyond technological advancements. This section discusses the broader societal implications of this transformative technology, focusing on key areas of concern and potential benefits for urban populations.

Numerous privacy and ethical issues are raised with the increasing number of edge devices and other sensors in urban areas [23, 35]. Edge computing's ability to collect data and surveillance might increase unprecedented levels of data skimming, which raises questions regarding the infringement of personal privacy. In addition, there is the potential for risk of algorithmic bias. If edge-based AI systems begin making critical urban decisions, they risk replicating or amplifying societal biases through entropic algorithms [36]. Indeed, the amount of data produced in smart cities raises questions regarding data ownership and control [9, 23]: who should own, control, and manage this vast quantity of data? Lastly, the widespread access to edge computing calls the notion of informed consent into question, as residents may not realize how their data is being collected or utilized.

This edge computing innovation can help or add to urban society's technological divide. Indeed, providing equitable access to smart city services is critical, and the promise of edge-enabled services reaching all residents despite socioeconomic status or tech literacy should be met [37]. While there was ample conversation about just how smart this technology will make future cities, the reality is that for a city to be truly "smart," digital literacy and education must be available to all so that citizens may access digital elements of life [38]. It will also be used for assistive devices for persons with disabilities and making smart cities accessible to all sides of the population.

Intelligent edge computing will be the cornerstone, reshaping how we govern cities and gauge citizen participation. The implementation of data-driven governance would make city analysis and the future design of solutions more responsive and

evidence-based to bring efficiency to services rendered by the cities [39]. Furthermore, edge computing can help develop a citizen engagement platform that paves the way for direct and immediate participation among the public in city planning and policy-making [40, 41]. However, this powerful data at the edge, rather than obscuring transparency and accountability, would be another layer of necessity to prevent sacrificing residents' liberties for large-scale industry cost optimization [23].

It will lead to great economic changes when intelligent edge computing is commonly applied in the fields of smart cities [42]. Some jobs are set to be automated because of the change in the job market, but this will result in new opportunities, specifically related to edge computing and smart city management. This will also help strengthen the ecosystem and create market opportunities for companies that introduce new business models and services, adding value to the local economy [43]. Moreover, economization as a consequence of resource optimization operated through edge computing will save costs for cities and citizens, drastically enhancing economic effectiveness [44].

Urban sustainability could benefit significantly from the use of intelligent edge computing. Smart grids and intelligent energy management systems will optimize resource consumption, resulting in enhanced power efficiency that minimizes carbon emissions produced by city habitats [44]. Edge-based systems will allow better environmental monitoring and response [45]: if air quality drops or waste management fails in an area, it can be automatically routed to deal with it. Edge computing also allows sustainable urban planning [11, 45], which plays a pivotal role for theorists to imagine more environmental data, pushing back that green line while making smarter development decisions based on this additional data.

While tackling these societal implications, a human-centric perspective on the emergence and operation of intelligent edge computing for smart cities is pivotal. This entails addressing the associated risks and challenges and using the technology proactively to make urban spaces more inclusive, equitable, and sustainable.

Intelligent edge computing for smart cities is entrenched intriguingly in the future. However, it will take continual dialog, policy design, and a commitment to ethical principles for this potential to materialize with the least harm done. This is a critical reminder for the path ahead—that we must ensure the future of our cities grows outward in ways that improve life for all its citizens, creating smart, equitable, inclusive, and resilient communities.

5. Challenges and opportunities

Implementing intelligent edge computing in smart cities presents a complex landscape filled with challenges to overcome and opportunities to seize [2]. This section wraps up this chapter with an overview of the critical technical, economic, and regulatory challenges and the potential for innovation and urban transformation.

Energy efficiency is one of the essential technical constraints of deploying edge computing in smart cities. Energy consumption becomes a concern, especially as the number of edge devices multiplies. Efficient, low-power hardware and algorithms become paramount for ensuring the sustainability of smart city infrastructures [46]. Additionally, with the speed at which cities are growing and changing, edge computing systems must be scalable. In addition to workhorse hardware devices, they will need software to efficiently orchestrate large numbers of distributed networks.

Incompatibility is still a significant issue [9]. With a plethora of edge devices, sensors, and platforms prevalent in smart cities, there is a need for standardized protocols that can maintain smooth communication as well as data interchange across systems [22]. At the same time, reliability and fault tolerance are essential because edge systems frequently underpin vital urban services. Creating systems that can heal themselves and maintain high availability is a continuous task in guaranteeing high service availability. However, security and privacy are also huge topics [9]. Top security measures are needed to ensure data processed by edge devices are safe from potential cyber threats but without the heavy performance drop that accompanies them.

Economic and regulatory challenges in implementing edge computing also contribute to the slow growth of smart cities [47]. Investment and ROI (return on investment) are among the key concerns. The costs of an initial investment in developing edge computing infrastructure can be significant, and cities and businesses need to determine how these projects will be funded while providing specific returns on those investments. Even when they do, the dynamic and fluid nature of edge computing tech innovation often leaves regulatory frameworks lacking in their wake. It is a complex ongoing challenge to strike the right balance in creating guidelines that ensure safety, privacy, and fair competition but at the same time do not choke-off or kill innovation.

Data governance is another critical issue, as cities must develop explicit rules about who owns the data, who can use it, and how [23]. It requires delicate management of legal and ethical concerns about how data is managed. Additionally, there is a digital talent gap in edge computing and AI, which is inhibiting implementation. The lack of experts with the requisite knowledge to effectively design, build, and operate smart city systems is a severe problem.

Despite these challenges, edge computing presents numerous opportunities for innovation and urban transformation [43]. One example is city services, which can be improved with edge computing, allowing for more responsive and efficient systems—from a smart healthcare offer to intelligent transportation solutions. It also gives cities the ability to make real-time decisions, such as emergency response situations or traffic management, by exploiting low latency processing capabilities within edge computing.

With edge computing, providing more context-aware services is easier, enabling a personalized experience of a smarter and quality life in the city [29]. In addition, the green aspect of these smart cities improves other subsystems, such as energy management demand response. Edge-enabled systems will enhance environmental sustainability by optimizing resource management and monitoring environmental conditions while offering new revenue streams. For instance, smart apps that provide unique insights about city activity. More tools allow citizens to participate in civic governance, monitoring or cooperation opportunities, and long-term urban development planning.

The edge computing economic growth potential cannot be ignored. It promotes innovation and entrepreneurship, spawns new employment markets, and induces economic growth in urban areas [4]. Similarly, urban planners can make more accurate and timely decisions by deploying edge systems with this rich real-time data to support better urban planning. Edge facilities also drive resiliency because they allow predictive maintenance and expedited repair for equipment failures, keeping cities running well in the face of outages. Lastly, edge-enabled systems also increase the empowerment of citizens by offering urban data delivered directly to residents and open decision-making processes that enhance their engagement with the system.

DOI: http://dx.doi.org/10.5772/intechopen.1007810

To leverage them for smart cities and benefit from intelligent edge computing, all these challenges must be addressed now, and future opportunities must be tapped wisely. Collaboration across sectors on technology, urban planning, policy-making, and community engagement will be essential. In the different dimensions of this interwoven scenario, seeing which path should be navigated through the advancement of technology and how it connects us with human-centered design is compulsory. We need to work toward building smart cities that leverage the benefits of emerging technologies like IoT and where people are at the center—citizens are our customers, always. Through carefully tackling the challenges and opportunities of intelligent edge computing, we can move closer to making urban environments not just "smart,: but also sustainable, inclusive yet able to be agile as the needs of each city community evolves.

Author details

Yu Chen[1]* and Ronghua Xu[2]

1 Binghamton University, Binghamton, New York, USA

2 Michigan Technological University, Houghton, Michigan, USA

*Address all correspondence to: ychen@binghamton.edu

IntechOpen

References

[1] Monzon A. Smart cities concept and challenges: Bases for the assessment of smart city projects. In: 2015 International Conference on Smart Cities and Green ICT Systems (SMARTGREENS). Piscataway, NJ, USA: IEEE; 2015. pp. 1-11

[2] Khan LU, Yaqoob I, Tran NH, Kazmi SA, Dang TN, Hong CS. Edge-computing-enabled smart cities: A comprehensive survey. IEEE Internet of Things Journal. 2020;7(10):10200-10232

[3] Chen N, Chen Y. Smart city surveillance at the network edge in the era of IoT: Opportunities and challenges. Smart Cities: Development and Governance Frameworks. 2018:153-176

[4] Jansson Å. Reaching for a sustainable, resilient urban future using the lens of ecosystem services. Ecological Economics. 2013;**86**:285-291

[5] Chen N, Chen Y. Anomalous vehicle recognition in smart urban traffic monitoring as an edge service. Future Internet. 2022;**14**(2):54

[6] Raharjana I. A systematic literature review of environmental concerns in smart-cities. In: IOP Conference Series: Earth and Environmental Science. Vol. 245. Philadelphia, PA, USA: IOP Publishing; 2019. p. 012031

[7] Xu R, Nikouei SY, Nagothu D, Fitwi A, Chen Y. Blendsps: A blockchain-enabled decentralized smart public safety system. Smart Cities. 2020;**3**(3): 928-951

[8] Mahadevappa P, Al-amri R, Alkawsi G, Alkahtani AA, Alghenaim MF, Alsamman M. Analyzing threats and attacks in edge data analytics within IoT environments. IoT. 2024; **5**(1):123-154

[9] Gharaibeh A, Salahuddin MA, Hussini SJ, Khreishah A, Khalil I, Guizani M, et al. Smart cities: A survey on data management, security, and enabling technologies. IEEE Communications Surveys & Tutorials. 2017;**19**(4):2456-2501

[10] Moubayed A, Shami A, Heidari P, Larabi A, Brunner R. Edge-enabled V2X service placement for intelligent transportation systems. IEEE Transactions on Mobile Computing. 2020;**20**(4):1380-1392

[11] Barone G, Buonomano A, Forzano C, Palombo A, Russo G. The role of energy communities in electricity grid balancing: A flexible tool for smart grid power distribution optimization. Renewable and Sustainable Energy Reviews. 2023;**187**:113742

[12] Ramírez-Moreno MA, Keshtkar S, Padilla-Reyes DA, Ramos-López E, García-Martínez M, Hernández-Luna MC, et al. Sensors for sustainable smart cities: A review. Applied Sciences. 2021; **11**(17):8198

[13] Tran TX, Hajisami A, Pandey P, Pompili D. Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges. IEEE Communications Magazine. 2017; **55**(4):54-61

[14] Dias I, Ruan L, Ranaweera C, Wong E. From 5G to beyond: Passive optical network and multi-access edge computing integration for latency-sensitive applications. Optical Fiber Technology. 2023;**75**:103191

[15] Park J, Yoo S. Evolution of the smart city: Three extensions to governance, sustainability, and decent urbanisation

from an ICT-based urban solution. International Journal of Urban Sciences. 2023;**27**(Suppl. 1):10-28

[16] Cayuela LJ. Smart city initiatives: Design and implementation analysis in three major global cities: Singapore, Copenhagen and San Francisco [bachelor thesis]. Barcelona, Spain: Pompeu Fabra University; 2024

[17] Myagmar-Ochir Y, Kim W. A survey of video surveillance systems in smart city. Electronics. 2023;**12**(17):3567

[18] Basch CH, LeBlanc M, Ethan D, Basch CE. Violence depicted in advertisements on LinkNYC kiosks in Manhattan, New York City. International Journal of Adolescent Medicine and Health. 2021;**33**(1): 20180033

[19] Planas-Carbonell A, Anguelovski I, Oscilowicz E, Pérez-del Pulgar C, Shokry G. From greening the climate-adaptive city to green climate gentrification? Civic perceptions of short-lived benefits and exclusionary protection in Boston, Philadelphia, Amsterdam and Barcelona. Urban Climate. 2023;**48**:101295

[20] Choi J, Kim HM. State-of-the-art of Korean smart cities: A critical review of the Sejong smart city plan. Smart Cities for Technological and Social Innovation. 2021:51-72

[21] Mishra P, Singh G. Energy management systems in sustainable smart cities based on the internet of energy: A technical review. Energies. 2023;**16**(19):6903

[22] Gracias JS, Parnell GS, Specking E, Pohl EA, Buchanan R. Smart cities—a structured literature review. Smart Cities. 2023;**6**(4):1719-1743

[23] König PD. Citizen-centered data governance in the smart city: From ethics to accountability. Sustainable Cities and Society. 2021;**75**:103308

[24] Popescul D, Murariu L, Radu LD, Georgescu MR. Digital co-creation in socially sustainable smart city projects: Lessons from the European Union and Canada. IEEE Access. 2024;**24**:71088-71108

[25] Hua H, Li Y, Wang T, Dong N, Li W, Cao J. Edge computing with artificial intelligence: A machine learning perspective. ACM Computing Surveys. 2023;**55**(9):1-35

[26] Iftikhar S, Gill SS, Song C, Xu M, Aslanpour MS, Toosi AN, et al. AI-based fog and edge computing: A systematic review, taxonomy and future directions. Internet of Things. 2023;**21**:100674

[27] Wen W, Demirbaga U, Singh A, Jindal A, Batth RS, Zhang P, et al. Health monitoring and diagnosis for geo-distributed edge ecosystem in smart city. IEEE Internet of Things Journal. 2023; **10**(21):18571-18578

[28] Xu R, Nagothu D, Chen Y. AR-Edge: Autonomous and Resilient Edge Computing Architecture for Smart Cities. London, UK: IntechOpen; 2024

[29] Chen M, Li W, Hao Y, Qian Y, Humar I. Edge cognitive computing based smart healthcare system. Future Generation Computer Systems. 2018;**86**: 403-411

[30] Muhammad G, Hossain MS. Emotion recognition for cognitive edge computing using deep learning. IEEE Internet of Things Journal. 2021;**8**(23): 16894-16901

[31] Miyake A. Quantum computation on the edge of a symmetry-protected

topological order. Physical Review Letters. 2010;**105**(4):040501

[32] Vitale A, Donati E, Germann R, Magno M. Neuromorphic edge computing for biomedical applications: Gesture classification using emg signals. IEEE Sensors Journal. 2022;**22**(20): 19490-19499

[33] Badidi E. Edge AI and blockchain for smart sustainable cities: Promise and potential. Sustainability. 2022;**14**(13): 7609

[34] Xu R, Ramachandran GS, Chen Y, Krishnamachari B. Blendsm-ddm: Blockchain-enabled secure microservices for decentralized data marketplaces. In: 2019 IEEE International Smart Cities Conference (ISC2). Vol. 2019. Piscataway, NJ, USA: IEEE. pp. 14-17

[35] Ahmad K, Maabreh M, Ghaly M, Khan K, Qadir J, Al-Fuqaha A. Developing future human-centered smart cities: Critical analysis of smart city security, data management, and Ethical challenges. Computer Science Review. 2022;**43**:100452

[36] Kontokosta CE, Hong B. Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. Sustainable Cities and Society. 2021;**64**:102503

[37] Du M, Zhang X, Mora L. Strategic planning for smart city development: Assessing spatial inequalities in the basic service provision of metropolitan cities. In: Sustainable Smart City Transitions. Routledge. London, UK: Taylor & Francis; 2022. pp. 113-132

[38] Carrasco-Sáez JL, Careaga Butter M, Badilla-Quintana MG. The new pyramid of needs for the digital citizen: A transition towards smart human cities. Sustainability. 2017;**9**(12):2258

[39] Kaluarachchi Y. Implementing data-driven smart city applications for future cities. Smart Cities. 2022;**5**(2):455-474

[40] Kopackova H, Komarkova J, Horak O. Enhancing the diffusion of e-participation tools in smart cities. Cities. 2022;**125**:103640

[41] Simonofski A, Asensio ES, De Smedt J, Snoeck M. Citizen participation in smart cities: Evaluation framework proposal. In: 2017 IEEE 19th Conference on Business Informatics (CBI). Vol. 1, 2017. Piscataway, NJ, USA: IEEE. pp. 227-236

[42] Kumar H, Singh MK, Gupta M, Madaan J. Moving towards smart cities: Solutions that lead to the smart city transformation framework. Technological Forecasting and Social Change. 2020;**153**:119281

[43] Ferraris A, Santoro G, Pellicelli AC. "Openness" of public governments in smart cities: Removing the barriers for innovation and entrepreneurship. International Entrepreneurship and Management Journal. 2020;**16**(4): 1259-1280

[44] Humayun M, Alsaqer MS, Jhanjhi N. Energy optimization for smart cities using IoT. Applied Artificial Intelligence. 2022;**36**(1):2037255

[45] Almalki FA, Alsamhi SH, Sahal R, Hassan J, Hawbani A, Rajput N, et al. Green IoT for eco-friendly and sustainable smart cities: Future directions and opportunities. Mobile Networks and Applications. 2023;**28**(1): 178-202

[46] Xie H, Huang R, Sun H, Han Z, Jiang M, Zhang D, et al. Wireless energy: Paving the way for smart cities and a greener future. Energy and Buildings. 2023;**297**:113469

[47] Richter MA, Hagenmaier M, Bandte O, Parida V, Wincent J. Smart cities, urban mobility and autonomous vehicles: How different cities needs different sustainable investment strategies. Technological Forecasting and Social Change. 2022;**184**:121857

Chapter 2

AR-Edge: Autonomous and Resilient Edge Computing Architecture for Smart Cities

Ronghua Xu, Deeraj Nagothu and Yu Chen

Abstract

With the rapid advancements in artificial intelligence (AI), the Internet of Things (IoT), and network communication technologies, recent years have witnessed a boom in smart cities that has dramatically changed human life and society. While many smart city applications rely on cloud servers, enabling comprehensive information fusion among users, smart devices, and service providers to provide diverse, intelligent applications, IoT networks' high dynamicity and heterogeneity also bring performance, security, and interoperability challenges to centralized service frameworks. This chapter introduces a novel Autonomous and Resilient Edge (AR-Edge) computing architecture, which integrates AI, software-defined network (SDN), and Blockchain technologies to enable next-generation edge computing networks. Thanks to capabilities in terms of logically centralized control, global network status, and programmable traffic rules, SDN allows for efficient edge resource coordination and optimization with the help of artificial intelligence methods, like large language models (LLM). In addition, a federated microchain fabric is utilized to ensure the security and resilience of edge networks in a decentralized manner. The AR-Edge aims to provide autonomous, secure, resilient edge networks for dynamic and complex IoT ecosystems. Finally, a preliminary proof-of-concept prototype of an intelligent transportation system (ITS) demonstrates the feasibility of applying AR-Edge in real-world scenarios.

Keywords: edge computing, internet of things (IoT), artificial intelligence (AI), security, Blockchain, software defined networks (SDN), smart cities, internet of vehicles (IoV), intelligent transportation systems (ITS)

1. Introduction

With the proliferation of the Internet of Things (IoTs) atop the fifth-generation and beyond (B5G) communication technology, tens of billions of physical devices with network connectivity allow for a big band of data. Thanks to fast advancements in artificial intelligence (AI) and big data technology, recent years have witnessed a boom in ubiquitous and sustainable applications deployed on powerful cloud servers to link heterogeneous IoT devices through the Internet. As a result, Smart Cities have become realistic, dramatically changing human life and society by constructing

intelligent, sustainable, and safe living environments [1, 2]. However, with the exponential increase of physical devices and continuous development of diverse smart applications, a conventional system architecture that is solo based on the cloud computing paradigm nevertheless encounters many problems in efficiently handling the massive IoT data, satisfying Quality of service (QoS), and providing security and interoperability guarantees demanded practical scenarios in high dynamic and distributed network environments.

By migrating partial computing, storage, and networking capabilities from centralized cloud servers to the network edge near the end users, edge computing has emerged as a promising paradigm to meet challenges on cloud-centric IoT applications, like reducing end-to-end latency and improving security and privacy [3]. With the breakthroughs in AI, especially for machine learning (ML) techniques, integration of ML with edge computing has become an inevitable trend in the data-driven intelligent applications brought by IoT. Unlike transitional AI applications relying entirely on cloud computing, edge intelligence [4] allows most of the distributed edge computing resources to achieve intelligent capabilities to support diverse user-defined services and applications in Smart Cities. Because edge computing acts as an intermediary service layer between physical devices and intelligent services on cloud servers, a hierarchical cloud-edge computing architecture is widely adopted by large-scale and complex IoT-based applications [5].

The smart applications atop edge intelligence and IoT networks have numerous benefits, such as ultra-low latency, flexibility, robustness, and privacy preservation [6]. Nonetheless, due to the inherently dynamic and distributed nature of traditional IoT-Edge networks, centralized service frameworks still face significant challenges in performance, scalability, security, and privacy. Thanks to crucial characteristics like decentralization, immutability, and tractability, Blockchain has demonstrated great potential to revolutionize various aspects of the economy and society. Integrating blockchain and edge computing into one system is promising to provide decentralized management and trustworthy services for dynamic and distributed IoT networks [7]. As an intelligence-enhancing enabler in 5G networked systems, the Software Defined Network (SDN) promises to apply ML techniques to the heterogeneous network infrastructure. Thanks to capabilities in terms of logically centralized control, global network status, and programmable traffic rules, SDN empowered with ML methods allows for efficiently organizing, managing, maintaining, and optimizing resources (e.g., computing, storage, and networking) within multi-dimensional and self-autonomous networked systems [8].

1.1 Main contributions

This chapter introduces a novel Autonomous and Resilient Edge computing network architecture called AR-Edge to enable next-generation edge computing networks (NextG Edge). AR-Edge is a secure-by-design system infrastructure that integrates Blockchain, SDN, edge computing, and AI/ML technology to meet the challenges of current IoT-based ecosystems. **Figure 1** demonstrates a conceptual architecture of AR-Edge for Smart Cities, which acts as a backbone framework to link heterogeneous IoT devices and complex smart application domains. The multiple pervasively deployed IoT devices (e.g., sensors, cameras, and smartphones), network devices, and databases construct a fundamental physical infrastructure that offers data and resources (computing, communication, and storage) for the essential functionality of applications.

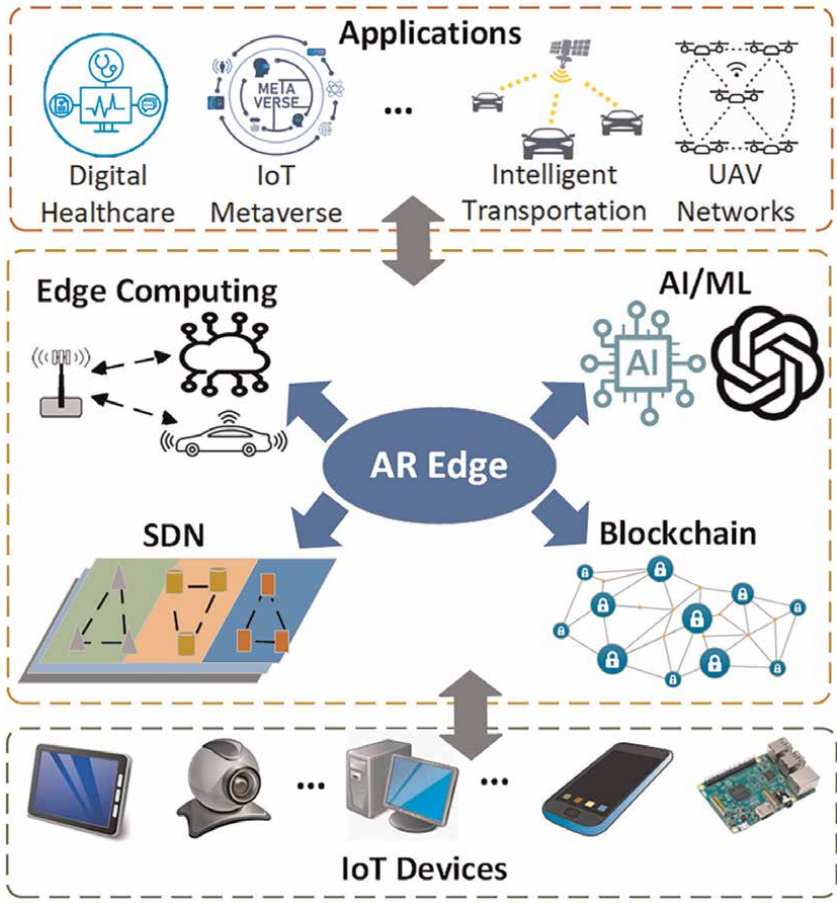

Figure 1.
Conceptual Architecture of AR-Edge for Smart Cities.

In AR-edge, SDN controllers work as actuators that focus on performance improvements by handling heterogeneous IoT networks and dynamic resource allocations. The centralized SDN controllers can have a global view of the network by monitoring real-time network state and operational configuration. Thus, the huge volume of collected data can facilitate the applications of AI/ML techniques to enhance the whole system. As a secure network fabric for AR-Edge, Blockchain introduces decentralization to mitigate single point failure caused by centralized management and service frameworks; therefore, it can ensure system resilience and availability under distributed network environments. In addition, decentralized security mechanisms atop Blockchain can protect IoT devices, edge computing platforms, SDN controllers, and even AI/ML models against cyber threats, such as Sybil attacks, unauthorized access to devices, modifying training data, privacy violations, and distributed denial of service (DDoS).

In AR-edge, edge computing technology is the cornerstone that benefits other core components by providing capabilities like computing, networking, and storage. By offloading ML training and reference tasks from central cloud servers to the distributed edge servers that are close to IoT devices, the response time of decision-making, data transmission delay, and network bandwidth cost can be dramatically reduced. In addition, edge computing also provides virtualization and softwarization platforms

that can manage physical devices and SDN controllers in efficient and standard manners. Moreover, edge computing platforms also offer resources for Blockchain networks, like computation required by miners or validators and storage for saving distributed ledgers. By leveraging powerful AI/ML methods, like large language models (LLM), AR-Edge relies on a global intelligent agent as the system brain to achieve a self-autonomous (self-adaptive, self-healing, and self-organizing) network infrastructure. Finally, AR-Edge aims to support various and multiple instances of IoT applications in Smart Cities (such as digital healthcare [9], IoT-enabled Metaverse [10], connected intelligent vehicles, and urban air mobility systems [11]).

The rest of this chapter is organized as follows. Section 2 describes fundamental concepts of Edge computing and AI integration. Section 3 introduces the general architecture and workflow of SDN and explains how SDN can be applied in AR-Edge to guarantee performance requirements. Section 4 explores emerging Blockchain technologies for AR-edge regarding security and privacy preservation. Section 5 presents a practical internet of vehicle example to illustrate how the proposed AR-Edge can be applied in real-world applications, like intelligent transportation systems (ITS) in smart cities. Finally, Section 6 provides a summary and discusses the open challenges and future directions.

2. AI-enabled edge intelligence for IoT

2.1 Overview of edge computing

The Internet of Things (IoT) concept was first introduced to the community in 1999 for supply chain management [12]. IoT enables a computer system to "feel" contextual information and respond with actions without human intervention. Therefore, it has greatly increasingly permeated human lives and become an essential enabling technology in Smart Cities. Due to powerful capabilities in computation and storage, the cloud computing paradigm is widely used to construct scalable service infrastructures that collect data from distributed IoT devices and provide services and applications for users in a centralized manner. In recent years, this world has witnessed the proliferation of IoT and cloud computing technology that connect users, applications, and devices through the Internet and continuously promotes a safe community and sustainable society [13]. With the continuous development of IoT devices, the massive growth of data, and various QoS requirements of applications, cloud computing-based infrastructures show many shortcomings, such as not providing a real-time response, privacy leakage, and high energy consumption [14].

As a new enabling technology that allows computation to be performed at the edge of the network, edge computing refers to computation and network resources along the path between data sources and cloud servers; downstream data represents cloud servers, and upstream data represent IoT devices [15]. Unlike the cloud computing model that collects data produced by IoT and then consumes them on the cloud servers, the edge computing model makes data aggregated and used (consumed by services) at the edge of the network. By migrating powerful computing and network capabilities and rich storage resources from cloud servers to the edge of the network, edge computing has various characteristics to serve intelligent services and critical applications based on distributed IoT networks. For example, the closer data source leads to low latency computing, the reduced network bandwidth usage achieves efficient energy consumption, transferred computing power to improve QoS and user experience in time-sensitive applications [16].

DOI: http://dx.doi.org/10.5772/intechopen.1005876

2.2 The convergence of AI and edge computing

Facilitated by the advancements of computing capabilities (e.g., hardware and software) and big data processing techniques, AI, especially for deep learning (DL), has achieved unprecedented success in various application domains, such as computer vision, autonomous driving, and natural language processing. The cloud data center uses unlimited storage to aggregate and save the massive amount of heterogeneous data transmitted from IoT devices and sensors. A cloud server provides ubiquitous on-demand computing capability for these computing-intensive smart services and applications. Due to the high overload of data transmission on network bandwidth, inherent latency constraints of network communication, and the risks of leaking private and sensitive information during data analysis, the cloud-centric framework is not suitable for time-critical and privacy-sensitive applications [17].

Thanks to many advanced features, such as low latency, reduced bandwidth consumption, energy efficiency, and privacy protection, edge computing promises to solve the above issues by moving the computational capability closer to the information-generation source. **Figure 2** demonstrates the convergence of AI and Edge computing from the perspective of a hierarchical IoT-Edge-Cloud paradigm. From a system architecture aspect, edge intelligence (EI) acts as a service infrastructure layer to connect heterogeneous physical devices and high-level tasks running on cloud servers. As a middleware layer between the IoT stratum and the edge stratum, the communication and virtualization layer leverages abstraction and softwarization techniques to manage virtual resources by mapping physical devices to virtual resources according to their capabilities, such as computation, connectivity, and storage. Therefore, the edge intelligence framework uses virtual resources to manage physical devices developed on different hardware platforms and connected via diverse communication protocols efficiently. The cloud server works as a global "brain" to store system-level knowledge databases and manage global trained model aggregation and reference. Thanks to comprehensive knowledge and advanced ML algorithms on cloud servers, intelligent collaboration can support resource orchestration and service adjustment on the EI layer to handle dynamic and complex system condition changes.

The critical function blocks in EI can be classified into four groups: network control, security, resource management, and AI/ML model. By applying ML methods to global network data, edge intelligence empowers learning capability to SDN controllers that provide intelligent network control to satisfy QoS given the complex network environments. As an essential function in network control, traffic classification allows SDN controllers to identify various traffic flows and perform fine-grained network management. Supervised and semi-supervised learning methods are widely used for traffic classification, which can be divided into elephant flow-aware traffic classification, application-aware traffic classification, and QoS-aware traffic classification [8]. Routing optimization is another important function in network control by enabling SDN controllers to modify flow tables in switches to achieve the optimal routing of traffic flows. By using a supervised learning algorithm called long short-term memory (LSTM) to estimate future network traffic, NeuRoute can calculate the optimal heuristic-like routing solutions in real-time [18]. Thanks to capabilities to solve decision-making problems, reinforcement learning (RL) algorithms are used to develop a distributed intelligent routing protocol, which allows SDN controllers to select optimal data transmission paths given the network status [19].

Both cryptographic mechanisms and Blockchains are widely used in edge intelligence to provide security and privacy-preserving guarantees. Encryption methods can

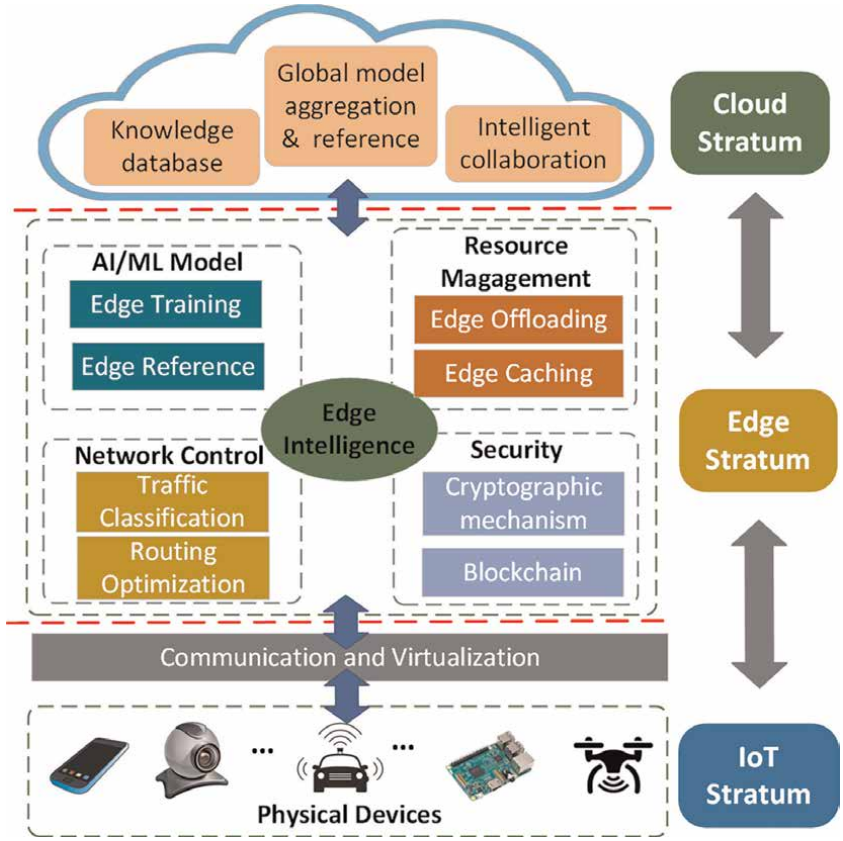

Figure 2.
The hierarchical framework of edge intelligence for IoT systems.

be used to ensure confidentiality during data transmission. Homomorphic encryption methods can especially protect private information in training and reference. In addition, asymmetric encryption and digital signature can provide identification authentication and access authorization for physical devices and virtual resources management. As a fundamental technology to ensure trust and decentralized network infrastructure, Blockchain can be integrated into EI to achieve dynamic resource orchestration and service re-adjustment for multi-domain IoT ecosystems [20].

As an extension of the cloud computing paradigm, edge computing offers infrastructure containing computing, storage, and network resources for edge training and edge reference. Edge training adopts a distributed learning architecture, like federated learning (FL), to learn the optimal values for all the weights and biases of DL models and identify hidden system patterns by training data sets stored at edge servers or IoT devices. For example, EI can provide a hierarchical FL framework to aggregate local training models into a global training model on the cloud server without directly exposing sensitive information of training data on devices [21]. Regarding high-quality EI service development, globally trained models and algorithms on cloud servers can be distributed at edge servers and devices to achieve high accuracy and low computation cost on testing instances.

Moreover, EI also provides resource management to support critical elements in intelligent applications and tasks, like training data collection, ML/DL model deployment, and computation provision. As EI leverages the edge computing paradigm to link distributed applications proximity to end-users and geographically scattered physical devices that record environmental information, edge caching techniques can collect the data generated from physical devices, such as networked cameras and sensors. The data are stored at reasonable places (edge servers or devices) and used for processing and analysis by intelligent algorithms to provide services for end-users [6]. For example, the raw video captured by cameras could be cached on edge servers for smart surveillance systems [22]. As an essential component of resource management in EI, edge offloading can leverage well-organized computing resources across the edge network to offer computing services for EI, like edge training and reference and network control.

3. SDN-based network control

3.1 SDN overview

The growing integration of edge devices has significantly increased the number of network management devices deployed to accommodate the traffic demand. Traditional network architecture relied on protocols such as Simple Network Management Protocol (SNMP), widely used to monitor network statistics and manually deploy network changes. Along with SNMP, NETCONF is another protocol more commonly used to automate the configuration of network devices. Although the protocols were available for remote configuration of networked devices and actively managing the traffic, the modern data throughput has significantly increased the bandwidth consumption closer to edge nodes and cloud data centers.

Integration of network devices from multiple manufacturers has increased the complexity of vendor-specific configurations, high cost, custom programming languages, and protocols specific to certain devices. For a constantly growing network connectivity, incompatibility among the networked devices could disable the network architecture functionality. To improve the interoperability among different vendors and create a network architecture capable of adapting to dynamic changes demanded by modern internet traffic, an efficient networking solution named Software Defined Networking (SDN) was devised.

The SDN has introduced a paradigm shift in the computer network architecture by decoupling the control plane and the data plane of the network devices. The data plane devices, such as the switches and the routers, are thereby responsible for network packet forwarding from interface to interface based on the instructions provided by the controller. The control plane maintains the bird's eye view of the network domain that the controller manages and provides user-instructed packet forwarding instructions to the data plane devices. The SDN consists of software-defined controllers that leverage Application Programming Interfaces (APIs) for managing the data plane hardware and leverage standardized protocols such as OpenFlow for their management [23]. The networked devices manufacturing industry has standardized the open communication channels among their devices using protocols such as OpenFlow, which enables remote changes to the device configurations using an SDN Controller. The integration of SDN has improved the interoperability among multiple vendor-networked devices and enabled dynamic changes to the link connectivity among the devices [24].

3.2 SDN architecture

As a result, the capability of separating the control plane and the data/forwarding plane has enabled SDN with dynamic and efficient programming of the network connections through a centralized controller hub [25]. To manage the devices in the forwarding plane, the SDN leverages interfaces through network connectivity. There are two dedicated application interfaces named Northbound and Southbound Interfaces. The Northbound Interface connects the SDN controller to the management plane where the network applications dedicated to the incoming network packets are running. Multiple network applications, such as firewall, tunneling, packet forwarding, load balancing, etc., are leveraged using the northbound interface. The Southbound interface connects the controller to the forwarding switches in the data plane to control the forwarding device, such as physical, wireless, optical, and virtual switches. The forwarding devices are controlled through supported network switches like open switches. However, it is not limited to that any forwarding device with supporting management protocols like OpenFlow, OVSDB, NetCONF, and SNMP can be controlled.

The SDN controller serves as the core of the control plane. Most SDN controllers cater services such as Topology, Inventory, Statistics, and Host tracking services. The topology service enables the discovery of the forwarding devices and their connectivity to other devices by using the Link Layer Discovery Protocol (LLDP) packets sent by the switch and observing the packet trajectory. The inventory service allows the SDN controller to track and record all SDN-enabled devices and their supported capabilities, like Openflow support. The statistics service reads counter information of the forwarding devices by using the flow table entries. Finally, the host tracking service discovers where the IP or MAC addresses are located on the network's topology.

Figure 3 represents the SDN architecture with individual components. With the SDN Controller established with access to the network applications on the Northbound interface and the forwarding devices on the Southbound interface, the controller can now manage the network traffic based on user specifications. For example, when a new host is connected to the switch in the data plane and tries to communicate with other devices in the network, the switch first consults the controller on instructions to handle the incoming packet from the new host. Based on the destination IP/

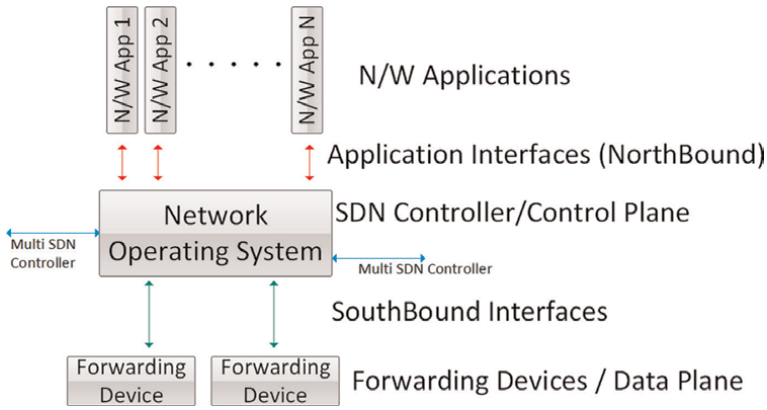

Figure 3.
SDN Architecture with North and Southbound Interfaces.

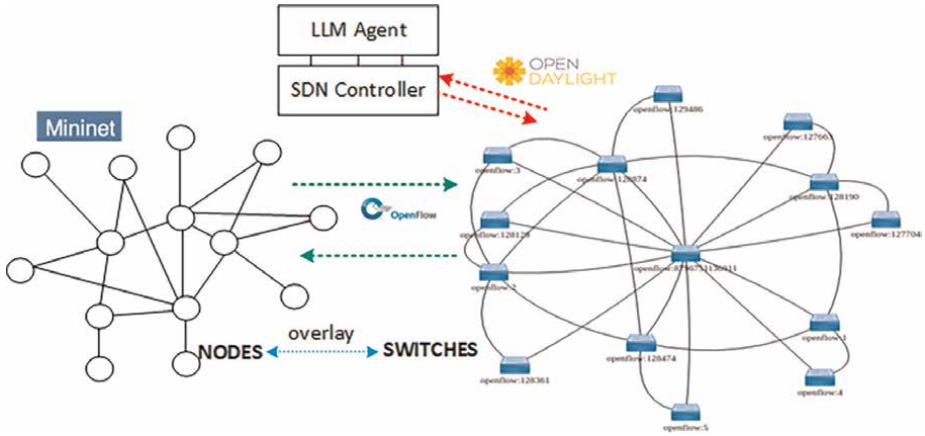

Figure 4.
SDN testbed initialized with Mininet and OpenDaylight SDN Controller.

MAC, the SDN Controller updates the switch with flow table entries, including instructions on handling such packets in the future with a timeout. With each switch flow table entries updated, the new host can communicate seamlessly with other nodes, and the controller maintains the inventory accordingly.

As the network scale gets larger, the resources in the SDN controller get limited. The centralized approach for SDN to manage the whole network could limit its optimal functionality for large-scale networks. However, multiple SDN controllers deployed to manage sectional networks allow for optimal functionality. The controllers update their inventory based on their East/Westbound interfaces, and the user can increase the scale of the network with unique or duplicate network applications catered by each controller. **Figure 4** represents a simulated network in a Mininet environment [26], and the OpenDaylight (ODL) SDN controller discovers the resulting switches [27]. Each node represents an open switch and can update its flow table entries based on the instructions received by the ODL Controller.

Leveraging network applications like routing optimization [18] for high-traffic network scenarios along with load balancing the SDN controller utilization with load balancing techniques, the network functionality is resilient to dynamic demands. Paired with LLM agents for supervised network traffic prediction and optimal route calculation, the autonomous SDN controller allows for an optimal edge network architecture.

4. Blockchain enhanced security

4.1 Blockchain overview

As the underlying technology of Bitcoin [28], *Blockchain* has demonstrated excellent capabilities to revolutionize traditional information systems based on the centralized system architecture. Essentially, Blockchain is a distributed ledger technology (DLT) atop decentralized Peer-to-Peer networks such that transactions and blocks are stored as a verifiable, append-only chained data storage. Blockchain leverages basic cryptographic primitives, such as cryptographic hash functions, asymmetric/

symmetric encryption, digital signature, and access control, to offer security guarantees. For example, cryptographic hash functions are widely used in consensus algorithms (e.g., Proof-of-Work and Proof-of-Stake) and committee election [29]. From a network architecture, Blockchain is an overlay network system that relies on Peer-to-Peer (P2P) network technology, like gossip communication [30] and distributed hash table (DHT) protocols [31, 32], to propagate control messages, transactions, and blocks without a centralized coordinator. The P2P network offers reliability and scalability for Blockchain data transmission in a large-scale distributed network environment.

As an essential function of a Blockchain system, consensus protocols allow all participants (miners or validators) to agree on the distributed ledger that maintains data integrity, consistency, and order of data across the distributed network without any third-party authority. The consensus in a distributed system aims to solve the Byzantine General Problem [33], which requires a single value agreement among different system parts given the failure of communication or conflicting information. Regarding various consensus protocols, Blockchain can be classified as permissionless blockchain (e.g., PoW and its variant) and permissioned blockchain (e.g., Practical Byzantine Fault Tolerant (PBFT) and its variant) [34]. Due to good scalability and global security in an open-access network environment, PoW has been used by public blockchain networks like Bitcoin and Ethereum. However, PoW requires high computation resources in mining blocks, such that it incurs unsustainable electrical energy consumption. PBFT [35] demonstrates better performances than PoW, such as low block confirmation latency, high transaction throughout, and less computation and energy consumption. However, it needs identity authentication and allows for limited network scalability regarding the number of validators during the consensus process.

By leveraging cryptographic primitives and secure computing mechanisms of Blockchain, *Smart Contract* can be used to develop self-enforcing and self-executing programs, which actuate the term of rules of a particular agreement or contract [36]. Smart contracts bring programmability to Blockchain by integrating business logic and user interfaces, such as offering complex operations and flexible services rather than solo cryptocurrency and cash-by-cash payment. Through publishing a set of application binary interfaces (ABIs), smart contracts act as autonomous trust agents between parties to fulfill predefined contract agreements under specific conditions [37]. Therefore, smart contracts can be used to implement decentralized applications (DApp) for services and applications under distributed and trust-less network environments.

4.2 Blockchain-based security solutions to edge computing

Thanks to the distributed nature of the edge network and the security guarantees of Blockchain, the integration of Blockchain and edge computing promotes decentralized, secure, and reliable EI services and applications.

Blockchain can be integrated into the EI system to ensure data transmission between physical devices, SDN controllers, and edge servers. For example, a blockchain-based distributed cloud architecture enables SDN-based fog nodes to interact with each other and through a blockchain network, and it can provide low-cost, secure, and on-demand access to the computing resources in a distributed edge computing network [38]. DistBlockNet [39] integrated Blockchain into a SDN-based edge computing network to update the flow rules table via SDN controllers, such switch devices can securely verify and validate flow rules table as downloading them.

Blockchain can be applied to the EI system to improve the capacity and security of data storage at edge computing networks. Due to the diverse formats and sizes of raw data generated by EI, Blockchain cannot directly store these data on the distributed ledger. Thus, hybrid on-chain and off-chain storage have been adopted by many Blockchain-based decentralized data storages for IoT applications [40]. All raw data are saved into off-chain storage, which is implemented by InterPlanetary File System (IPFS) [41] or Swarm [42]. While the metadata and reference of raw data are stored in transactions committed on Blockchain for verification during data accessing and sharing. Thus, Blockchain and decentralized storage can provide reliable, traceable, and tamper-proof data services without sacrificing security and privacy preservation.

5. Case study: towards security and resilience of intelligent transportation system

With the development of communication, network, and mobile computing technology, vehicles equipped with smart devices, such as wireless sensors, onboard computers, GPS antennas, cameras, radar, and so on, can collect and process large amounts of context-aware data while enabling information exchange between vehicles [43]. Thanks to advancements in vehicular communication and self-driving technology, Internet of Vehicles (IoV) becomes realistic through seamless interconnection among smart vehicles, roadside infrastructure, pedestrians, transportation service providers, and intelligent traffic management systems. By enabling a comprehensive information exchange platform between highly connected smart vehicles and heterogeneous vehicular services, an intelligent transportation system (ITS) leverages AI and big data techniques to provide diverse intelligent and safe vehicular applications, such as enhanced pedestrian and driving safety, efficient traffic planning, smart parking, and entertainment services [44].

The ITS aims for a seamless, connected, and ubiquitous service platform that supports large volume data collecting, transacting, and sharing among participants such as vehicles, pedestrians, and service providers. Meanwhile, existing ITS systems that rely on cloud-based storage and management incur new concerns on performance, scalability, interoperability, security, and privacy. First, ever-increasing interconnected vehicles, along with a massive amount of vehicular data, introduce scalability issues in the centralized ITS services framework. In addition, future ITS also considers interoperability as sharing data and resources with a wide range of service providers, such as original equipment manufacturers (OEM), insurance companies, and transportation departments. Because ITS uses advanced AI/ML algorithms and models to provide diverse smart applications, edge computing nodes allow for caching data from vehicles and offloading tasks from cloud servers to improve QoS. Moreover, the centralized frameworks adopted by ITS are prone to performance bottlenecks and single points of failure under highly dynamic and distributed network environments. Therefore, it is necessary to rethink the system architecture for next-generation IoV networks and intelligent service platforms.

5.1 Design rationale and system architecture

To address the aforementioned issues in current IoV ecosystems, a novel system architecture is introduced by integrating the AR-Edge framework with multi-domain IoV networks to ensure the security and resilience of ITS. **Figure 5** demonstrates a

system architecture of interconnected vehicular service networks based on a crossroad scenario. The whole IoV network consists of four independent and fragmented vehicle networks. Each vehicle network adopts an AR-Edge framework to perform domain-specific network control, resource orchestration, intelligent services deployment, and security and privacy enforcement. As a permissioned network, each vehicle network relies on system administration services deployed on edge servers to manage all registered entities within a vehicle network. As **Figure 5** shows, these entities could be vehicles, cameras, traffic lights, roadside unit (RSU), smart devices used by pedestrians, communication infrastructure, edge servers, charging stations, etc. Each registered entity within a vehicle is assigned a unique identifier (UID) by a trust system administrator.

ITS ecosystems use IoV networks and computing technology to enable information and data sharing among participants and achieve a self-organized network. The information interaction between vehicles and other entities refers to vehicle-to-everything (V2X) models, which include vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-roadside unit (V2R), vehicle-to-grid (V2G), and vehicle-to-pedestrian (V2P) [43]. Each vehicle uses onboard control units and deployed sensors (e.g., inertial measurement unit (IMU), radar and camera, etc.) to collect and process car operating states and environmental information. After joining an IoV network, the vehicle can use an embedded telematics control unit (TCU) to offer wireless communication to and from vehicles and other entities. V2V allows vehicles to broadcast useful information to each other, such as emergency braking, collision detection,

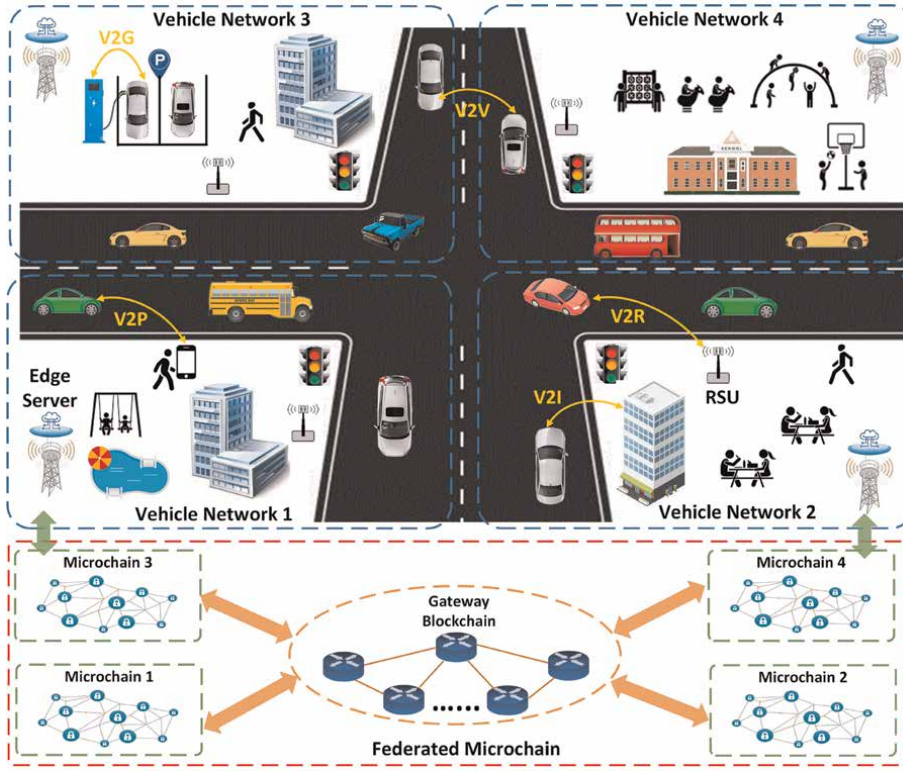

Figure 5.
The system overview of AR-Edge-enabled IoV networks.

and road traffic conditions, thereby improving user driving safety and travel efficiency [43].

In an IoV network, RSUs are placed along the roadside to support V2R and V2I for information exchange between vehicles and other network infrastructures. The RSUs are edge computing nodes that provide storage to save vehicle data and traffic information within a vehicle network. In addition, deploying network control functions on RSUs can improve network resource allocation and interoperability between vehicle communication and heterogeneous service and application networks. Furthermore, RSUs can perform local decision-making based on real-time traffic conditions to offer optimal vehicle travel routines within the covered service range. By collecting vehicle data and traffic information, the edge server of a vehicle network acts as the "smart brain" that maintains the global traffic status and performs global traffic strategy. The cloud server processes computation-intensive LLM training tasks on multidimensional transportation system data (vehicle data, road traffic information, network states, and services deployment.) Given individual vehicle network status and QoS requirements, a fine-tuned model can be deployed on the edge server that performs offloading intelligent tasks to optimize transportation service provision and ensure public safety for drivers and pedestrians.

As a decentralized and trust-free network infrastructure for ITS, a federated microchains framework [20] is integrated into IoV ecosystems to ensure reliable and secure data sharing and resource allocation across multiple vehicle networks. The existing Blockchain-based IoV solutions rely on monolithic Blockchains such that they cannot handle the Blockchain trilemma [45]. However, the hierarchy of the federated microchains promises to make trade-offs in decentralization, efficiency, scalability, and security by applying blockchain to dynamic and heterogeneous IoV networks. As the bottom of **Figure 5** shows, each vehicle network relies on a microchain to record data and transactions within the network. At the same time, a global gateway Blockchain links multiple fragmented microchains. In each microchain, a periodically randomly selected committee executes a lightweight consensus protocol, such as PBFT or Delegated Proof-of-Stake (DPoS), to verify and store data on a private distributed ledger. In addition, a microchain leverages a hybrid storage scheme by combining an on-chain ledger and an off-chain distributed database to guarantee the security and privacy of sensitive data within a vehicle network.

From a global security aspect, delegating nodes selected from vehicle networks construct a global gateway blockchain to guarantee scalability and interoperability for multi-domain IoV networks. The gateway Blockchain relies on a PoW consensus protocol to maintain a publicly distributed ledger. For multi-domain operations, raw data are saved on permissioned microchains, while references or metadata of original data are saved into checkpoint blocks that are finalized on the gateway Blockchain. Therefore, gateway nodes can utilize inter-microchain protocols that rely on global checkpoint blocks to ensure auditability, immutability, and provenance for cross-microchain transactions.

5.2 Prototype implementation and evaluation

To verify the feasibility of integrating AR-Edge with IoV networks, a proof-of-concept prototype is implemented with Python and tested on a virtual network environment by using Mininet [26]. The network topology of the testbed consists of two virtual IoV networks and one virtual gateway blockchain network, all deployed on an HPC. All virtual networks are connected through a remote controller deployed on a

	HPC-Workstation	Desktop
CPU	2.2GHz, Intel(R) Gold 5520R (96 cores)	3.4GHz, Core (TM) i7-2600 K (8 cores)
Memory	512GB DDR3	16GB DDR3
Storage	4 TB HHD	500GB HHD

Table 1.
Configuration of experimental devices.

desktop. **Table 1** describes devices used for the experimental study. This case study evaluates how a microchain federation can improve the security and interoperability of multidomain IoV ecosystems. Therefore, a private Ethereum Blockchain [46] is set up on one virtual gateway blockchain network, while the private Tendermint blockchain [47] is configured on two other virtual IoV networks. Multiple virtual hosts are created for each virtual network to simulate entities of IoV networks or blockchain gateway nodes, and each node is assigned one cup core.

The test cases are developed to evaluate the latency and throughput of processing transactions under test scenarios: query inter-ledger transactions and commit inter-ledger transactions given varying gateway nodes and system transaction throughput. The number of gateway nodes ranges from 4 to 20. The system transaction throughput Th_S is defined as "users send" transactions per second (tps) during query and commit data operations. The processing transaction throughput Th_P is denoted as the actual "system can process" transactions per second (tps) during inter-ledger operations. We conducted 50 Monte Carlo test runs for each case scenario and used the average of the results for evaluation.

First, we set four gateway nodes for each virtual network and evaluate latency and throughput by increasing Th_S from 20 to 1000 tps. **Table 2** shows the total processing latency of transactions as scaling up Th_S. Given a fixed number of gateway nodes, the end-to-end delays incurred by query operations are almost linear to Th_S. Because they are dominated by system capabilities, such as each host's computing power, network link bandwidth, etc. The underlying Blockchain's properties, like block confirmation time, have significant impacts on the latency of committing transactions on the distributed ledger. As a result, the end-to-end delays caused by commit operations are almost stable when $Th_S \leq 200$tps. **Figure 6** plots trends of processing transaction throughput Th_P as scaling up Th_S. When $Th_S \leq 200$ tps, Th_P is almost linear to Th_S due to the stable latency of committing inter-ledger transactions. However, system capability becomes the performance bottleneck as $Th_S \geq 200$ tps such that Th_P becomes saturated. Th_P of query inter-ledger transactions demonstrates almost stable because the capacity of gateway nodes mainly influences them.

To evaluate how the number of gateway nodes influences performance, we fixed $Th_S = 1000$ for both query and commit scenarios while increasing the number of gateway nodes from four to 20. **Table 3** shows the total processing latency of

TPS	20	50	100	200	500	1000
Query inter-ledger transactions (second)	0.1	0.2	0.5	1.0	2.3	4.5
Commit inter-ledger transactions (second)	1.6	1.6	1.6	1.6	3.0	5.5

Table 2.
The latency as scaling up system transactions.

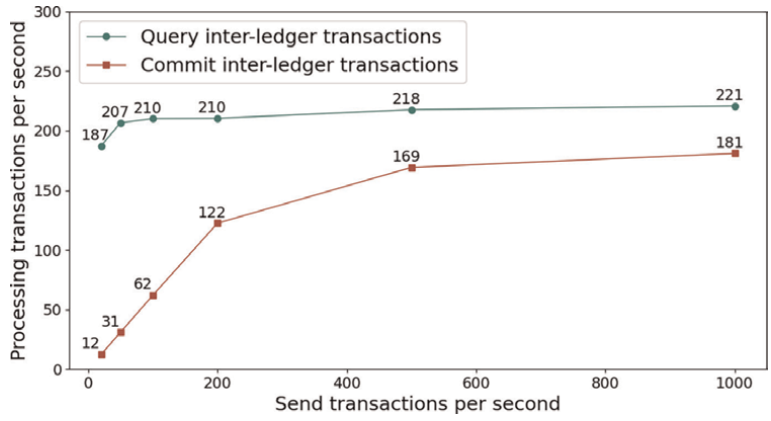

Figure 6.
Comparison of the throughput as scaling up transactions.

Number of nodes	4	6	8	10	12	16	20
Query inter-ledger transactions (second)	1.8	1.2	0.9	0.7	0.6	0.5	0.4
Commit inter-ledger transactions (second)	2.9	1.6	1.6	1.6	1.6	1.6	1.6

Table 3.
The latency as scaling up gateway nodes.

transactions as scaling up gateway nodes. AR-Edge can introduce resource allocation to dynamically configure gateway nodes and network control to coordinate transactions within the gateway blockchain network efficiently. Therefore, adding more gateway nodes to distributed service overload can dramatically reduce the total latency of processing a large volume of transactions in a short time period, especially for query operations. **Figure 7** demonstrates trends of processing transaction throughput Th_P as scaling up gateway nodes. Th_P of query inter-ledger transactions are almost linear to increase the number of gateway nodes. However, the performance

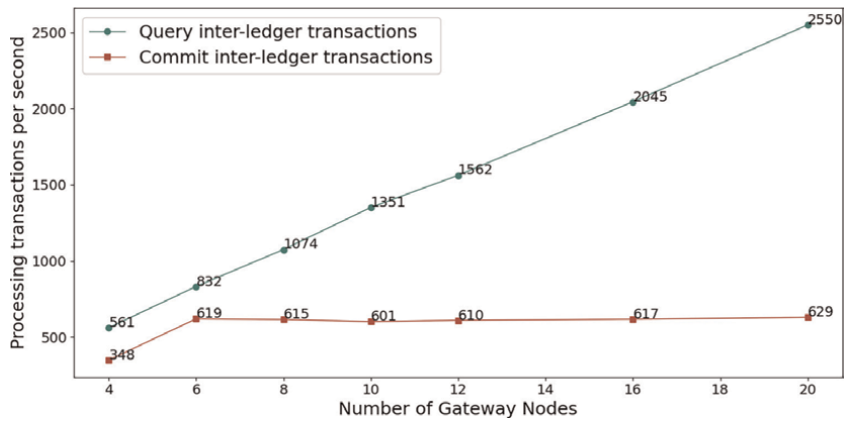

Figure 7.
Comparison of the throughput as scaling up gateway nodes.

of inter-ledger transactions is mainly dominated by the characteristics of gateway Blockchain using a PoW consensus protocol. Thus, solo increasing the number of gateway nodes cannot reduce processing latency or improve Th_P when the system is handling a large volume of transactions within a short time (e.g., $Th_S = 1000$). Applying a lightweight consensus protocol with short transaction committed time and high processing throughput into the gateway blockchain is a promising solution to improve system performance, especially for committing data on the distributed ledger.

6. Conclusions

This chapter introduces an AR-Edge that adopts AI/ML, edge computing, SDN, and Blockchain as enabling technologies to construct the next edge computing networks for diverse applications in smart cities. Given a comprehensive overview of AR-Edge architecture, we present a hierarchical framework of edge intelligence atop IoT ecosystems. We explain how core enabling technologies collaboratively promote an intelligent, secure, and resilient system architecture for dynamic and complex IoT scenarios. Finally, an ITS based on multiple IoV networks is demonstrated as a case study. The prototype analysis shows that AR-Edge is a promising services and applications framework in IoT ecosystems.

However, questions and open issues remain unanswered, such as designing AR-Edge in real-world scenarios and validating system performance and security features. One challenge is the gateway blockchain architecture design that ensures efficiency, security, and interoperability for cross-micro chain operations, especially for cases that integrate cryptocurrency networks (e.g., bitcoin and Ethereum) to support decentralized financial services. Another challenge is investigating AI/ML techniques to achieve intelligent SDN controllers that optimize resource allocations and enhance system security with proactive prediction and mitigation of cyber threats. Our ongoing efforts are developing an SDN-based edge intelligence framework that can be applied to practical applications, such as ITS atop a scalable IoV network and urban air mobility (UAM) system in smart cities.

Acknowledgements

This work is supported by the Institute of Computing and Cybersystems (ICC) at Michigan Technology University via Rapid Seeding award 24-0560-P0001.

Abbreviations

AI	artificial intelligence
ABI	application binary interfaces
API	application programming interface
DApp	decentralized applications
DDoS	distributed denial-of-service
DHT	distributed Hash table
DL	deep learning
EI	edge intelligence
FL	federated

IMU	inertial measurement unit
IoT	Internet of Things
IoV	internet of vehicles
IPFS	interplanetary file system
ITS	intelligent transportation systems
LLDP	link layer discovery protocol
LLM	large language models
LSTM	long short-term memory
ML	machine learning
ODL	OpenDaylight
OEM	original equipment manufacturer
DPoS	delegated proof-of-stake
PoW	proof-of-work
P2P	peer-to-peer
QoS	quality-of-service
RSU	roadside unit
RL	reinforcement learning
SDN	software defined network
SNMP	simple network management protocol
TCU	telematics control unit
UAV	unmanned air vehicle
V2G	vehicle-to-grid
V2I	vehicle-to-infrastructure
V2P	vehicle-to-pedestrian
V2R	vehicle-to-roadside unit
V2V	vehicle-to-vehicle
V2X	vehicle-to-everything

Author details

Ronghua Xu[1], Deeraj Nagothu[2] and Yu Chen[3]*

1 Department of Applied Computing, Michigan Technological University, Houghton, USA

2 Intelligent Fusion Technology, Inc., Germantown, USA

3 Department of Electrical and Computer Engineering, Binghamton University, Binghamton, New York, USA

*Address all correspondence to: ychen@binghamton.edu

References

[1] Xu R, Lin X, Dong Q, Chen Y. Constructing trustworthy and safe communities on a blockchain-enabled social credits system. In: Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. New York, NY, USA: ACM; 2018. pp. 449-453

[2] Xu R, Nikouei SY, Nagothu D, Fitwi A, Chen Y. Blendsps: A blockchain-enabled decentralized smart public safety system. Smart Cities. 2020;**3**(3): 928-951

[3] Yu W, Liang F, He X, Hatcher WG, Lu C, Lin J, et al. A survey on the edge computing for the Internet of Things. IEEE Access. 2017;**6**:6900-6919

[4] Zhou Z, Chen X, Li E, Zeng L, Luo K, Zhang J. Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE. 2019;**107**(8):1738-1762

[5] Hamdan S, Ayyash M, Almajali S. Edge-computing architectures for internet of things applications: A survey. Sensors. 2020;**20**(22):6441

[6] Xu D, Li T, Li Y, Su X, Tarkoma S, Jiang T, et al. Edge intelligence: Empowering intelligence to the edge of network. Proceedings of the IEEE. 2021; **109**(11):1778-1837

[7] Yang R, Yu FR, Si P, Yang Z, Zhang Y. Integrated blockchain and edge computing systems: A survey, some research issues and challenges. IEEE Communications Surveys & Tutorials. 2019;**21**(2):1508-1532

[8] Xie J, Yu FR, Huang T, Xie R, Liu J, Wang C, et al. A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges. IEEE Communications Surveys & Tutorials. 2018;**21**(1):393-430

[9] Qu Q, Xu R, Sun H, Chen Y, Sarkar S, Ray I. A digital healthcare service architecture for seniors safety monitoring in metaverse. In: 2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom). New York, NY, USA: IEEE; 2023. pp. 86-93

[10] Qu Q, Hatami M, Xu R, Nagothu D, Chen Y, Li XE, et al. Microverse: A Task-Oriented Edge-Scale Metaverse. Future Internet. 2024;**16**(2):60

[11] Xu R, Wei S, Chen Y, Chen G, Pham K. LightMAN: A lightweight microchained fabric for assurance-and resilience-oriented urban air mobility networks. Drones. 2022;**6**(12):421

[12] Ashton K et al. That 'Internet of Things' thing. RFID Journal. 2009;**22**(7): 97-114

[13] Nagothu D, Xu R, Nikouei SY, Chen Y. Smart surveillance for public safety enabled by edge computing. In: Edge Computing: Models, Technologies and Applications. 2020. pp. 409-433

[14] Cao K, Liu Y, Meng G, Sun Q. An overview on edge computing research. IEEE Access. 2020;**8**:85714-85728

[15] Shi W, Cao J, Zhang Q, Li Y, Xu L. Edge computing: Vision and challenges. IEEE Internet of Things Journal. 2016; **3**(5):637-646

[16] Kong X, Wu Y, Wang H, Xia F. Edge computing for internet of everything: A survey. IEEE Internet of Things Journal. 2022;**9**(23):23472-23485

[17] Chang Z, Liu S, Xiong X, Cai Z, Tu G. A survey of recent advances in edge-computing-powered artificial intelligence of things. IEEE Internet of Things Journal. 2021;**8**(18):13849-13875

[18] Azzouni A, Boutaba R, Pujolle G. NeuRoute: Predictive dynamic routing for software-defined networks. In: 2017 13th International Conference on Network and Service Management (CNSM). New York, NY, USA: IEEE; 2017. pp. 1-6

[19] Sendra S, Rego A, Lloret J, Jimenez JM, Romero O. Including artificial intelligence in a routing protocol using software defined networks. In: 2017 IEEE International Conference on Communications Workshops (ICC Workshops). New York, NY, USA: IEEE; 2017. pp. 670-674

[20] Xu R, Chen Y, Li X, Blasch E. A secure dynamic edge resource federation architecture for cross-domain IoT systems. In: 2022 International Conference on Computer Communications and Networks (ICCCN). New York, NY, USA: IEEE; 2022. pp. 1-7

[21] Xu R, Chen Y. μDFL: A secure microchained decentralized federated learning fabric atop IoT networks. IEEE Transactions on Network and Service Management. 2022;**19**(3):2677-2688

[22] Xu R, Nagothu D, Chen Y. Decentralized video input authentication as an edge service for smart cities. IEEE Consumer Electronics Magazine. 2021; **10**(6):76-82

[23] Shalimov A, Zuikov D, Zimarina D, Pashkov V, Smeliansky R. Advanced study of SDN/OpenFlow controllers. In: Proceedings of the Ninth Central and Eastern European Software Engineering Conference in Russia. New York, NY, USA: ACM; 2013. pp. 1-6

[24] Nunez A, Ayoka J, Islam MZ, Ruiz P. A brief overview of software-defined networking. arXiv preprint arXiv:230200165. 2023

[25] Raghunath K, Krishnan P. Towards a secure SDN architecture. In: 2018 Ninth International Conference on Computing, Communication and Networking Technologies (ICCCNT). New York, NY, USA: IEEE; 2018. pp. 1-7

[26] Mininet. An instant virtual internet on your desktop (or the PC). Available from: http://mininet.org/ [Accessed: January, 2024]

[27] OpenDaylight. [Accessed: January, 2024]. Available from: https://opendaylight.org/

[28] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. Bitcoin. 2008;**4** (2):15. Available from: https://bitcoin.org/bitcoin.pdf

[29] Xu R, Nagothu D, Chen Y. ECOM: Epoch randomness-based consensus committee configuration for IoT Blockchains. In: Principles and Practice of Blockchains. New York, NY, USA: Springer; 2022. pp. 135-154

[30] Jenkins K, Hopkinson K, Birman K. A gossip protocol for subgroup multicast. In: Proceedings 21st International Conference on Distributed Computing Systems Workshops. New York, NY, USA: IEEE; 2001. pp. 25-30

[31] Stoica I, Morris R, Karger D, Kaashoek MF, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications. ACM SIGCOMM Computer Communication Review. 2001;**31**(4):149-160

[32] Maymounkov P, Mazieres D. Kademlia: A peer-to-peer information

system based on the XOR metric. In: International Workshop on Peer-to-Peer Systems. New York, NY, USA: Springer; 2002. pp. 53-65

[33] Lamport L, Shostak R, Pease M. The byzantine generals problem. In: Concurrency: The Works of Leslie Lamport. New York, NY, USA: ACM; 2019. pp. 203-226

[34] Xu R, Chen Y, Blasch E. Microchain: A light hierarchical consensus protocol for IoT systems. In: Blockchain Applications in IoT Ecosystem. New York, NY, USA: Springer; 2020. pp. 129-149

[35] Castro M, Liskov B, et al. Practical byzantine fault tolerance. In: OsDI. Vol. 99. New York, NY, USA: ACM; 1999. pp. 173-186

[36] Hewa T, Ylianttila M, Liyanage M. Survey on blockchain based smart contracts: Applications, opportunities and challenges. Journal of Network and Computer Applications. 2021;**177**:102857

[37] Xu R, Chen Y, Chen G, Blasch E. SAUSA: Securing access, usage, and storage of 3D point CloudData by a blockchain-based authentication network. Future Internet. 2022;**14**(12): 354

[38] Sharma PK, Chen MY, Park JH. A software defined FOG node based distributed blockchain cloud architecture for IoT. IEEE Access. 2017; **6**:115-124

[39] Sharma PK, Singh S, Jeong YS, Park JH. Distblocknet: A distributed blockchains-based secure SDN architecture for IoT networks. IEEE Communications Magazine. 2017;**55**(9): 78-85

[40] Xu R, Chen Y. Fed-DDM: A federated ledgers based framework for hierarchical decentralized data marketplaces. In: 2021 International Conference on Computer Communications and Networks (ICCCN). New York, NY, USA: IEEE; 2021. pp. 1-8

[41] Benet J. IPFS-content addressed, versioned, P2P file system. arXiv preprint arXiv:14073561. 2014

[42] Swarm. [Accessed: January, 2024]. [Online]. Available from: https://ethersphere.github.io/swarm-home/

[43] Ji B, Zhang X, Mumtaz S, Han C, Li C, Wen H, et al. Survey on the internet of vehicles: Network architectures and applications. IEEE Communications Standards Magazine. 2020;**4**(1):34-41

[44] Mollah MB, Zhao J, Niyato D, Guan YL, Yuen C, Sun S, et al. Blockchain for the internet of vehicles towards intelligent transportation systems: A survey. IEEE Internet of Things Journal. 2020;**8**(6):4157-4185

[45] Zhou Q, Huang H, Zheng Z, Bian J. Solutions to scalability of blockchain: A survey. IEEE Access. 2020;**8**: 16440-16455

[46] Go-ethereum. [Accessed: January, 2024]. Available from: https://ethereum.github.io/go-ethereum/

[47] Tendermint core. [Accessed: January, 2024]. Available from: https://docs.tendermint.com/master/

Chapter 3

An Effective and Efficient Computation Architecture for Edge Computing Devices on IoMT-Based Deep Belief Networks

Dinesh Kumar Jayaraman Rajendiran, Balaji Venkatesalu Ramasamy, Joby Titus T, Karthi Samiyampalayam Palanisamy and Visvesvaran Chandramohan

Abstract

The Internet of Medical Things (IoMT) is one of the growing and emerging technologies in recent trends. Faster computation is the major requirement of any edge computing device. Edge computing systems require Effective Computation Blocks (ECB) to store and process signals between users and the cloud. The time taken for trans-receiving and processing of the signal should be minimal, which is mentioned as latency. The ECB assures high-end power transmissions, especially in autonomous vehicles, robotic surgery, diagnosis, and medicine distributions. The ECB architecture is based on highly effective computation. The computation is independent of internet connectivity and therefore the major suspect is uncertainty. This work focuses on the development of sustainable approximation adder for edge devices of IoMT. This architecture performance is measured at the deep learning architectures which are familiar at the edge devices of cloud computing. In cases of low internet, the computing devices are slower, which causes all devices and applications to go down the track. By implementing the proposed adder (PAXA) at the edge, computing gets around the dependencies by locating data that is closer to the possibility, which speeds up applications and improves their availability and also in the applications where it requires high speed and low-power availability.

Keywords: effective computation block (ECB), edge computing, internet of medical things (IoMT), low power, deep believe network (DBN)

1. Introduction

1.1 Overview of edge computing devices and IoMT

The convergence of deep belief networks (DBNs) and edge computing devices under the Internet of Medical Things (IoMT) brings both opportunities and

challenges in the fast-changing field of healthcare technology. The deep belief networks, which are well-known for their capacity to collect intricate medical data and derive significant insights, need strong computing units that are optimized for edge situations. For several reasons, an efficient and productive computing unit is essential for edge computing devices in IoMT-based DBNs. By maximizing computational efficiency and resource consumption, it first tackles the intrinsic limits of edge devices, such as constrained processor power, memory, and energy resources [1]. Second, by lowering the latency involved in data transmission to centralized servers or cloud platforms, a computation unit allows real-time inference and decision-making at the edge and artificial intelligent (AI) devices. This capability is particularly critical in time-sensitive medical applications where prompt diagnosis and intervention can significantly impact patient outcomes. Furthermore, an optimized computation unit facilitates the seamless integration of DBNs with IoMT devices, ensuring interoperability, data privacy, and security compliance [2]. The use of actionable insights from diverse data sources including wearable sensors, medical imaging equipment, and electronic health records, gives healthcare providers more power. IoMT-based DBNs open the door to personalized medicine, remote patient monitoring, and proactive healthcare interventions by utilizing the power of an effective and efficient computation unit.

This ultimately improves the quality of care delivery and patient experience in the digital age of healthcare. The integration of deep belief networks (DBNs) into the Internet of Medical Things (IoMT) ecosystem presents a promising avenue for advancing healthcare analytics and decision-making processes [3]. Deep learning networks (DBNs) are a subclass of algorithms that are particularly good at identifying intricate patterns and characteristics in medical data, such as diagnostic imaging and patient records. But in the Internet of Medical Things (IoMT), their implementation on edge computing devices needs a specialized processing unit. It strikes a balance between effectiveness and efficiency [4]. To support the computational demands of deep learning algorithms, an efficient computation unit for edge computing devices in IoMT-based DBNs is preferred. It must guarantee low latency, energy efficiency, and real-time inference capabilities. This calls for the integration of hardware accelerators designed specifically for DBNs, such as field-programmable gate arrays (FPGAs) or specialized neural processing units (NPUs). Additionally, the computing unit ought to interact with IoMT devices seamlessly to facilitate effective preprocessing, edge model deployment, and data transfer [5]. Through the utilization of such processing units, IoMT-based DBNs can open up new possibilities for decentralized, real-time medical analytics, allowing for prompt insights and actions to enhance patient outcomes and the effectiveness of healthcare delivery. **Figure 1** illustrates the architecture of IoMT general representation.

1.2 Server model of IoMT

The Internet of Medical Things (IoMT) refers to the integration of biosensors, medical devices, healthcare systems, and applications through interconnected technologies, enabling the exchange of health-related data. IoMT leverages the principles of the broader Internet of Things (IoT) to transform the healthcare landscape by enhancing patient care, improving efficiency, and providing valuable insights to healthcare providers. As IoMT continues to evolve, the integration of edge computing plays a pivotal role in addressing challenges related to data processing, latency, and security.

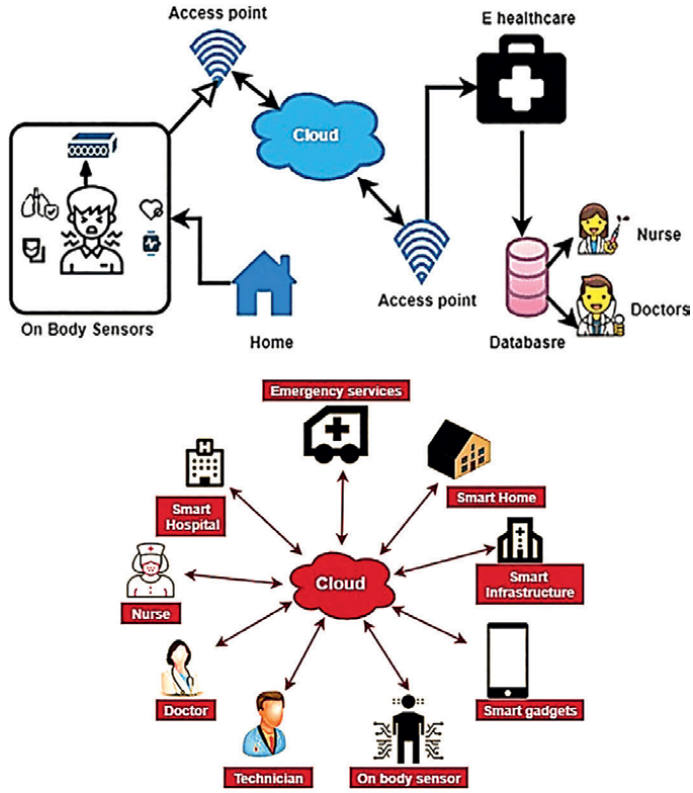

Figure 1.
IoMT architecture – general representations.

The major components of IoMT are medical devices and connectivity. Medical devices consist of various sensors, which can extract the features of the real-time physical signal from the human body. The received signals are processed by the converter to provide input to the processor, which is the core of all devices. This process is undergone by distinct elements such as wearable, implantable, and monitoring devices. The next component of IoMT is the cloud platform which is used for connectivity [6]. The known platform for connectivity is the cloud, where every device can be connected. It helps to analyze and realize the physical variations obtained from the human body or healthcare-connected systems. The modes of transmission of Wi-Fi are Bluetooth, wireless sensor nodes (WSN), and body area network (BAN). The dataset is extracted from the available data source or cloud-connected data models [7]. The cloud-connected model processes the real-time signal and accesses the available information as the hyper-tuned parameters [7]. This parameter accesses the functionality of the body organs and is used for agile diagnosis. **Figure 2** illustrates user server interaction for accessing the data through IoMT protocols whose architecture is similar to **Figure 1** representation [8].

1.3 Impact of IoMT and associated computation models

The integration of edge computing in the Internet of Medical Things (IoMT) represents a paradigm shift in healthcare technology, addressing critical challenges associated with data processing, latency, and the need for real-time decision-making.

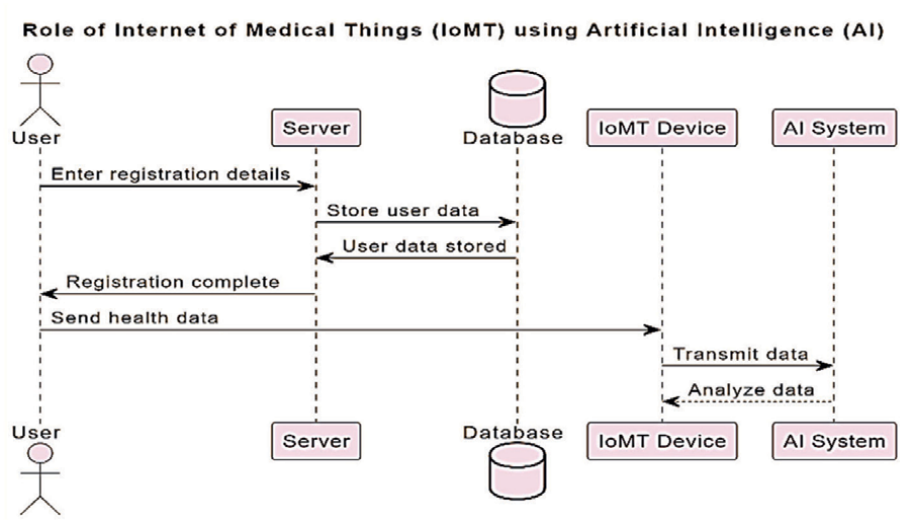

Figure 2.
User–server interaction for accessing the data via IoMT.

This section explores the key aspects of edge computing in IoMT, highlighting its significance and impact on healthcare [9]. Edge computing involves processing data closer to the source of data generation rather than relying solely on centralized cloud servers. In IoMT, it means bringing computational tasks to the edge of the network, allowing for immediate processing of data or near the point of collection. The notable significance of edge devices on IoMT is continuous monitoring, rapid response, and real-time analytics. Edge computing ensures that critical data is processed locally, reducing latency and enabling timely interventions. **Figure 3** illustrates the end user applications related to the IoMT architecture [10]. It brings the advantages of delay reduction (latency reduced), real-time analysis, and optimization of bandwidth with improved data privacy.

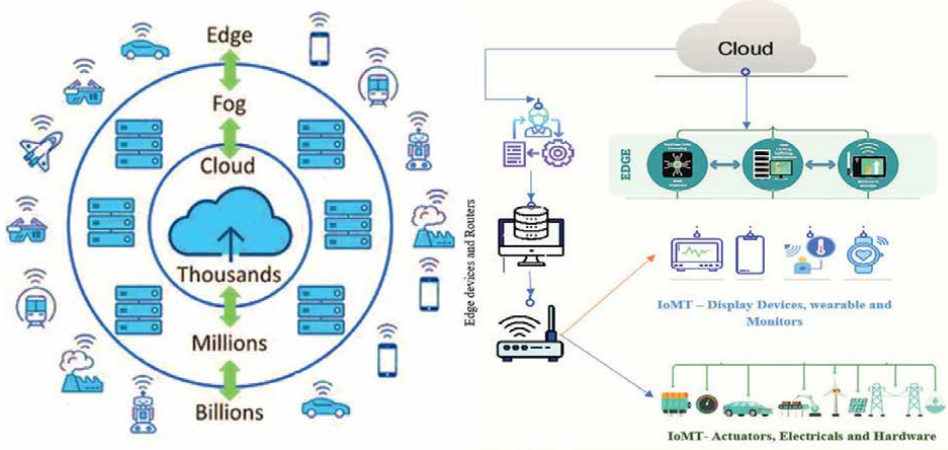

Figure 3.
IoMT – architecture and end-user applications.

1.4 Architecture and implementation: edge devices and gateways

The integration of IoT and edge computing involves the deployment of edge devices and gateways strategically within the healthcare infrastructure. These devices act as intermediaries between IoMT devices and centralized cloud servers, handling data processing tasks locally. Fog computing, an extension of edge computing, is particularly relevant in IoMT [11]. It involves distributing computing resources across the network, providing hierarchical levels of data processing based on application requirements, while edge computing enhances data privacy. The substantial volume of data generated by IoMT devices necessitates real-time analysis to enhance the performance and service quality of IoMT applications. From streamlined data collection to efficient diagnosis, IoMT has automated healthcare processes, transforming conventional systems into smart ones. IoMT sensors gather vital patient data, facilitating real-time sharing with healthcare providers and caregivers. IoMT's contribution in collecting and analyzing patient data autonomously, and storing results in the cloud, allows for instant insights and statistics. This technology ensures healthcare services are accessible anytime, anywhere, enabling caregivers to manage multiple patients simultaneously and significantly improving traditional systems [12]. The burgeoning IoMT technology offers growth opportunities, especially in the healthcare sector, with advancements in communication technologies and protocols.

1.5 Understanding deep belief networks (DBNs)

Deep belief networks are a type of deep learning model composed of multiple layers of latent variables (typically Restricted Boltzmann Machines or RBMs) that form a probabilistic generative model. DBNs are trained using unsupervised learning, usually with a method called Contrastive Divergence or Gibbs Sampling. DBNs require significant computational resources for training and inference, especially as the model size increases. Efficiency is critical to ensure real-time processing and minimize energy consumption.

The hardware accelerators such as GPUs, TPUs, or custom application specific integrated circuits (ASICs) are tailored for deep learning inference. These accelerators are optimized for parallel processing, which is beneficial for the matrix operations involved in DBNs. Quantization techniques are used to reduce the precision of weights and activations, thus reducing memory and computational requirements without significant loss in accuracy [13]. The utilization of on-chip memory helps to reduce the need for external memory access, which can significantly reduce energy consumption and latency. The validation and testing are done through simulation, emulation, and real-world testing in IoMT environments. Various metrics such as inference speed, energy efficiency, and resource utilization are also considered to check the performance.

The integration of the computation unit in edge computing devices is commonly used in IoMT applications, such as wearable health monitors, medical imaging devices, or implantable medical devices. Selecting the appropriate processor architecture is crucial. Advanced risc machine (ARM) Cortex-A series processors are commonly used in edge devices due to their low power consumption and scalability. Alternatively, RISC-V architectures offer flexibility and customization options. These processors are widely used in a variety of devices, including smartphones, tablets, embedded systems, and increasingly edge computing devices [14]. Cortex-A processors are based on the ARMv7-A or ARMv8-A instruction set architectures, which

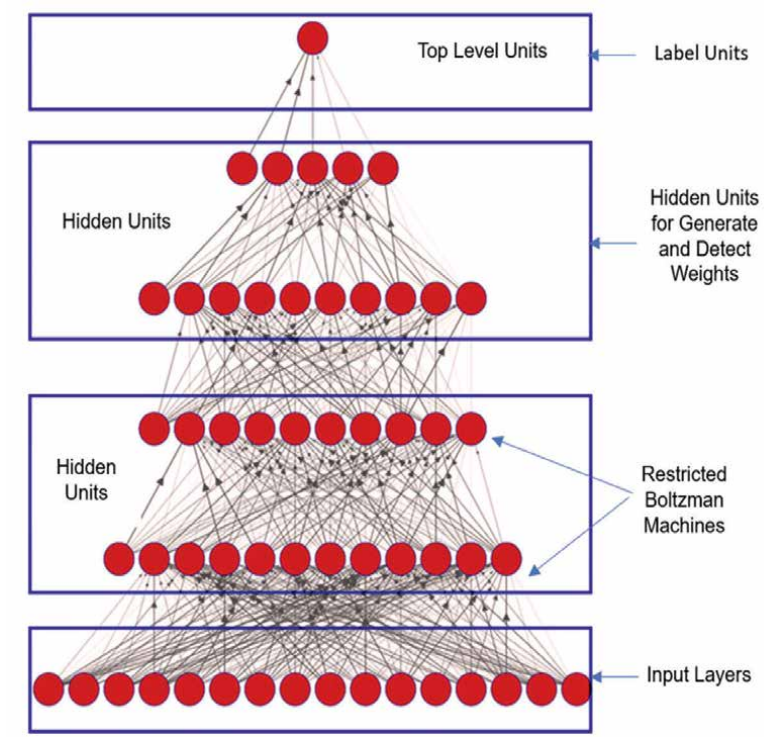

Figure 4.
Deep belief networks – general representations.

provide support for 32- and 64-bit processing, respectively. ARMv8-A introduces features like AArch64 (64-bit execution state), improved SIMD (single instruction, multiple data) support, and enhanced security features. Many Cortex-A series processors are available in multicore configurations, allowing for parallel execution of tasks to improve overall system performance.

These processors can be implemented in symmetric multiprocessing (SMP) or heterogeneous computing setups, where different types of cores (e.g., high-performance and low-power cores) are combined to optimize power efficiency. **Figure 4**, illustrates the deep belief networks general representations. Cortex-A series processors are often integrated into system-on-chip (SoC) solutions along with other components such as graphics processing units (GPUs), accelerators, memory controllers, and connectivity interfaces [15]. This integration enables highly integrated and compact designs for edge computing devices.

2. Existing methods and background survey

The Author Samie et al. proposed the survey methods that provide a comprehensive overview of edge computing in IoT, covering key components, architectures, and future directions [16]. It discusses the role of computing units at the edge of the network and analyzes various challenges and opportunities in this domain. Wang et al. presented a comprehensive survey of edge intelligence technologies, focusing on their role in enabling intelligent IoT applications. It explores the integration of computing

units at the edge to support real-time processing and decision-making and discusses emerging trends and research directions [2–6, 11]. Sarkar et al. discussed the architecture, applications, and advantages of edge computing in IoT environments [17]. It highlights the role of computing units deployed at the edge of the network to reduce latency, enhance scalability, and improve the efficiency of IoT deployments. Guo et al. explain the overview of edge computing technologies and their applications in IoT [18]. It discusses the role of computing units in edge environments and examines various architectural designs, communication protocols, and optimization techniques for efficient edge computing. Yu et al. present a comprehensive survey of edge computing principles, technologies, architectures, and applications [11]. It discusses the role of computing units in edge environments and examines emerging trends, challenges, and opportunities for advancing edge computing in IoT [19]. Amr Soliman et al. explore the design of energy-efficient approximate adders specifically for deep learning accelerators. Although it targets a different application domain, the principles of energy efficiency and approximation may apply to edge IoT devices where power constraints are crucial. Xang et al. insights into various approximate adder architectures and techniques [6]. While not focusing explicitly on edge IoT, it discusses the trade-offs between accuracy and power consumption, which are important considerations for resource-constrained IoT devices. Xue Wang et al. express an approximation adder design approach that uses the ABC tool to evaluate the approximate adder's area and re-encodes the adder outputs using a genetic algorithm under the average error (AE) constraint.

The 2-, 3-, and 4-bit approximation adders are implemented using this technique [20]. Nithysree et al. explain that approximation circuits are utilized to improve speed and decrease design compilation. In multi-media applications, the necessary data can be gathered from slightly erroneous outputs; therefore, precise results are not necessary [21]. The approximation adders could be employed in certain applications. This study employs and tests a variety of approximation full adders at various bit locations within a ripple carry adder. The performance and error percentage of each approximate adder are compared. **Table 1** lists the existing papers and corresponding observations [23].

3. Proposed architecture: approximation and computation blocks

The deep learning networks (DLN), especially in neural networks, are designed for attention mechanisms in Deep Learning (DL). The computational efficiency is crucial due to the vast amount of matrix operations involved. Approximate computing techniques, such as approximate adders, are introduced to reduce computational complexity, memory requirements, and power consumption without significantly compromising the model's performance [24]. Role of approximate adder in deep attention belief neural network is involved in (i) reducing computational complexity, (ii) memory and energy efficiency, (iii) performance trade-off, (iv) scalability, and (v) hardware cost-effective design. The DLN uses many hidden layers to extract complex feature extraction. The extraction process is capable of learning and executing the task from the provided dataset. Even the multi-layer perceptron (MLP) is capable of handling vast databases by the hidden network. However, the proficiency of MLP is affected by the temporal data and its non-sophisticated functional methods. This drawback pushes the convolution and recurrent neural networks to perform the classification and data extractions, which are cited as CNN and RNN in the following

Reference	Year	Methodology	Key Focus	Observation
Rosa et al. [15]	2023	Dynamic Voltage and Frequency Scaling (DVFS)	Accuracy, throughput	Focuses on adaptable approximation strategies for varying workloads
Manohar et al. [19]	2023	Error-Tolerant Computing	Reliability, energy efficiency	Investigate error-resilient designs suitable for resource-constrained environments
Tridi et al. [17]	2023	Input Sparsity Exploitation	Accuracy, latency	Explores trade-offs between accuracy and speed in edge computing tasks
Liu & Li et al. [18]	2022	Error-Resilient Design	Power, area, accuracy	Proposed error-resilient design with selective accuracy degradation for improved power and area efficiency
Kim et al. [22]	2021	Voltage Scaling	Energy efficiency	Demonstrated voltage scaling technique to achieve 15% energy savings in edge devices
Tripati et al. [13]	2021	Voltage Scaling	Power consumption, delay	Investigate low-power approximation in edge devices
Daniel et al. [14]	2021	Bit Prediction	Accuracy, area, energy	Proposes a novel technique for reducing area overhead in edge devices
Smith et al. [8]	2020	Error-Tolerant Logic	Area, power, latency	Proposed error-tolerant adder reduces power by 30% with minor accuracy loss
Chen et al. [9]	2019	Stochastic Computing	Area, power, throughput	Utilized stochastic computing to reduce area by 20% with minimal power overhead
Zhang & Wang et al. [11]	2018	Approximate Computation	Accuracy, throughput	Implemented approximate adders with adjustable precision to trade-off between accuracy and throughput

Table 1.
Existing models and methods used for approximate computing in IoMT.

discussions. CNN has a better architecture than MLP in terms of processing the data and effective spatial fitting and filtering of data. The competence of CNN makes it suitable for processing data from the image by entrapping the local information. This makes the CNN to differ from MLP [18].

The CNN has many convolution layers, which are used to extract the pixel information from the images and process by (i) convolution, (ii) pooling activation, and (iii) fully connected layers. The convolution layer is used to extract spatial information through its architecture. **Figure 5** represents the CNN architecture representations for 512*512-pixel images in the DL environment. The convolution network is formulated by the adders and multipliers. This layer performs the convolution of 1D or 2D signals by considering one as stationary and another as time-delayed varying signal. The sum of the product of these two signals is marked as the convolution layer. Hence, this layer plays a significant role in transferring the spatial data into the higher-level data to identify patterns and features. The next layer is pooling and activation. It is performed as space reduction by extracting the most suitable information

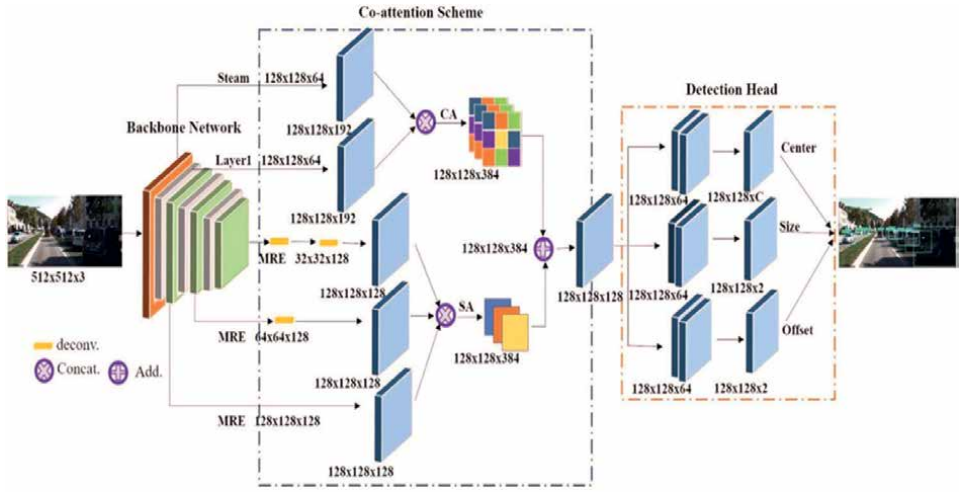

Figure 5.
*Representation of CNN based on 512*512 image models on DL environment.*

from the convoluted signals. The activation unit is the second half of this layer, which could help in shifting the layers of hidden network from non-linear to linear and vice versa by redeeming the originated information. The fully connected layer has an authority on decision making by combining responses from the previous layers. The CNN has a similar layer to CNN. However, it depends on the back looping, whereas another traditional method follows the forward looping. Therefore, RNN is helpful in analyzing the sequence of data, audio, and signals by maintaining the proper template memory management section. The layers of RNN are similar to CNN, except for the presence of a recurrent layer for observing the layer information for long-term dependent data. The recurrent layers are formulated by long-short-term memory (LSTM), which processes the sequence of data and measures the spatial temporary information. This LSTM depends on the time and sequence provided to the model. It reduces the vanishing gradient problem by the gate or cell structure and long-term dependency [25]. The cell consists of (i) cell state – for the chain of memory structure and holds the spatial information for a longer duration. (ii) Forgot cell – for removing the redundant information and discarding the unclear pixels obtained from the existing hidden layer. (iii) Input cell – This is used to update the next sequence and new data obtained from the previous layers. (iv) The authority cell of output is used to generate the classified results and allow the next data to be processed. Also, this unit discards the existing information and updates the new data by intimating the model. In spite of all considerations, the RNN suffers by vanishing issues, which are gradient in pattern.

This issue pushes the RNN to modify the architecture by providing the complex reduced optimized model for effective classifications. Each model has the limitation of processing the spatial non-linear data from the provided dataset. The generative adversarial network (GAN) is another model for processing the sequence of data by the synthetic data process. The GAN may depend on the windows and weight mechanism applied to its nodes and processes in the graphical structures. Here, the variable autoencoder (VAE) is used to represent the data generated from the GAN generative networks. Therefore, an effective computation unit is needed at the edge devices to process the data via DLN models. The recently developed model is the

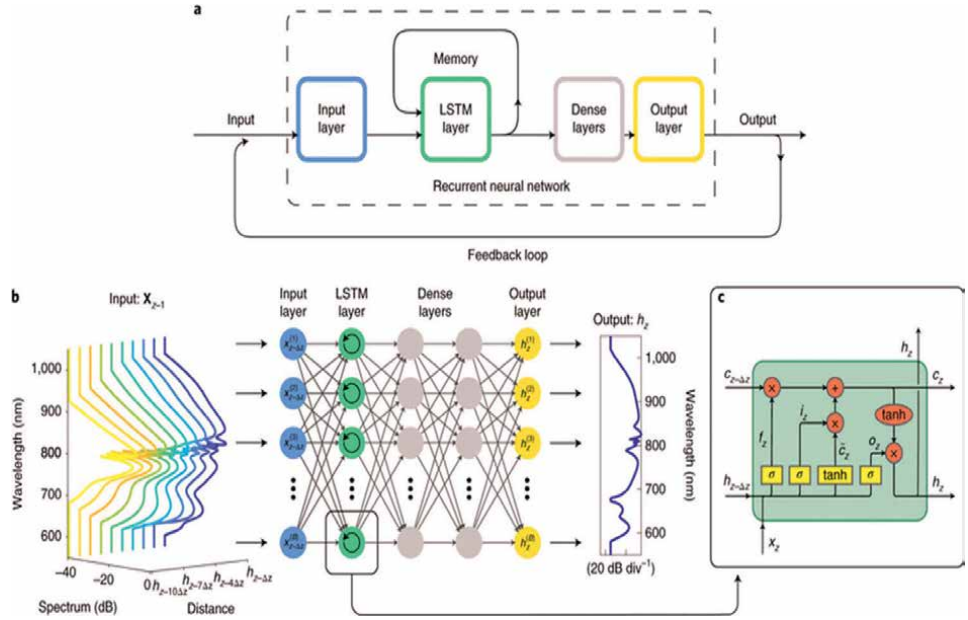

Figure 6.
Computation blocks of LSTM and RNN for edge computing devices in IoMT.

attention network of transformer (ATNT) model, which is designed by combining the significance of CNN, RNN, and GAN. **Figure 6** visualizes the computation blocks involved in LSTM and RNN architectures, which have been used in edge computing devices recently. The major layer of the ATNT is available as bi-variations. Such as self and multi-head. The self ATNT (SATNT) is available within the model of distinct node representation in the graph, and multi-ATNT (MATNT) is vice versa [26]. The architecture of SATNT is made by generating the sequence of elements and the impacts on the same sequence to observe all the variations. This model performance depends on the layers of computation and the corresponding score. This is given by the weighted matrix at the SoftMax layers and the interaction with the existing models on the same sequence. This way of analysis helps the model to agile response for the particular sequence of variations. The MATNT is the multiple combination of SATNT. Here, each is dependent and executed in parallel. Therefore, the data at the input layers are processed by the different sectional heads of SATNT and provincially joined at the output node. This enables the user of MATNT is capable of capturing various headed data and analyzing the data independently. The GAN model under the ATNT recreates the transformer architecture similar to LSTM and enables quick response for spatial non-linear data processing. The encoding methods in this architecture enable the element to identify the relevant and actual positions for classifying the data patterns. The LSTM method has the ability to handle the data of long dependency [17]. The LSTM includes the hidden state of fixed size, which reduces the loss in the information of dependent long sequences. In addition to this, the encoding technique is used at inherent position for the data and carries the maximum information through the data dependency sequence. **Figure 6** illustrates the LSTM model for the encoding and decoding sequence of long-dependent data. As mentioned earlier, the data has been processed by the embedding block, where the positional encoding is performed.

The proposed approximate adder (PAXA) is the multi-head attention layer that accepts the embedded information and processes via the adder–normalization (add and norm) block. There are two sections of Add–Norm block for encoding and three sections for decoding. The multi-head layer is used for LSTM with an attention network for classification. The encoding and decoding are outperformed by the approximated adder design proposed here. In this work, the adder is replaced by the approximate adders. The approximate adders have improved in terms of lower power, low delay, and agile response [27]. **Figure 7** demonstrates the computation blocks of the deep belief network layer with an attention mechanism based on multiply accumulator (MAC). The LSTM mechanism is sectioned by three functions, namely value (VA), key (K), and the query (QU). These values are generated by the input after being processed by the multiple processes of inputs. These components are required for the computations at the LSTM, especially at the attenuation level. The QU is meant for the contextual value of the data-dependent values, and the key represents the variations of QU for the allotted weight. The VA represents the attention value score where the QU, K, and VA are compared to generate the final value on classification [28]. The attention value is used for the classification which is done by the pair of consecutive VA.

The VA is processed by SoftMax to ensure the binary value variations 1 and 0. The normalized binary value is used here. Hence, the performance is better than that of other methods. Also, the binary is encoded easily. The weights are assigned to each bot and multiplied to produce the variable weights and the contextual encoding. These values are used to learn the different aspects of the data in which the criteria of the information are handled for the contextual long data. Proximate adder, which are functionality for edge devices, focuses on optimizing arithmetic operations to meet the stringent constraints of power, area, and performance in resource-constrained environments [29].

Edge devices, such as smartphones, IoT devices, and wearable gadgets, often operate under limited computational resources and battery power, making energy-efficient arithmetic units crucial for their functionality. In the context of edge computing, approximate adders offer a promising solution by trading off a certain degree

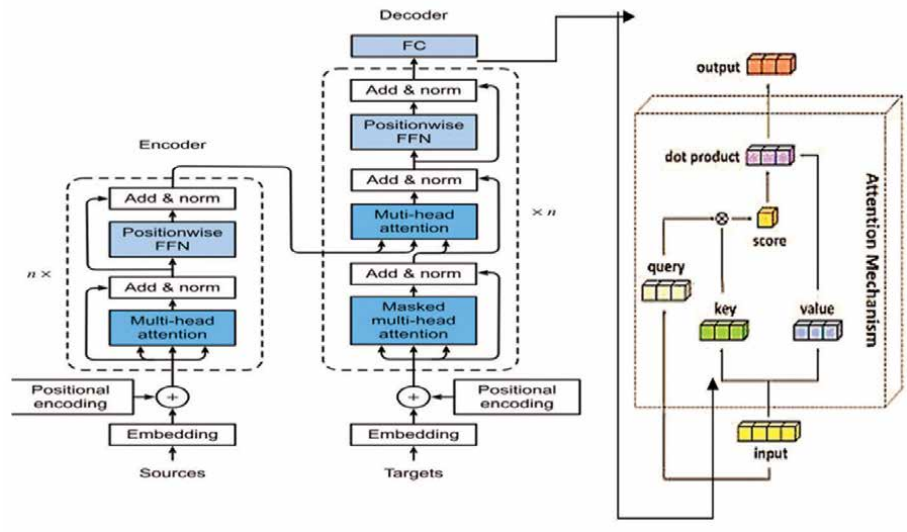

Figure 7.
Computation blocks of deep believe network layer with attention mechanism based on MAC.

of numerical accuracy for reduced hardware complexity and energy consumption. By employing simplified carry propagation schemes, probabilistic methods, or truncated arithmetic techniques, approximate adders can significantly reduce the number of transistors, gates, and power-hungry operations compared to conventional adders. This hardware efficiency is particularly beneficial for edge devices that often perform repetitive arithmetic operations in applications like sensor data processing, real-time analytics, and machine learning inference. **Figure 8**(a) depicts the general flow model of approximation method and evaluation steps and (b) illustrates LUT-based implementation approximate adder. In these scenarios, the slight error introduced by approximate adders is usually tolerable and can be mitigated through algorithmic or software-level error compensation techniques [30].

Furthermore, approximate adders enable edge devices to execute arithmetic operations more quickly, facilitating faster data processing and response times, which is critical for real-time applications. The reduced power consumption also extends the battery life of edge devices, enhancing their usability and reliability in remote or mobile environments where recharging may not be readily available. The PAXA

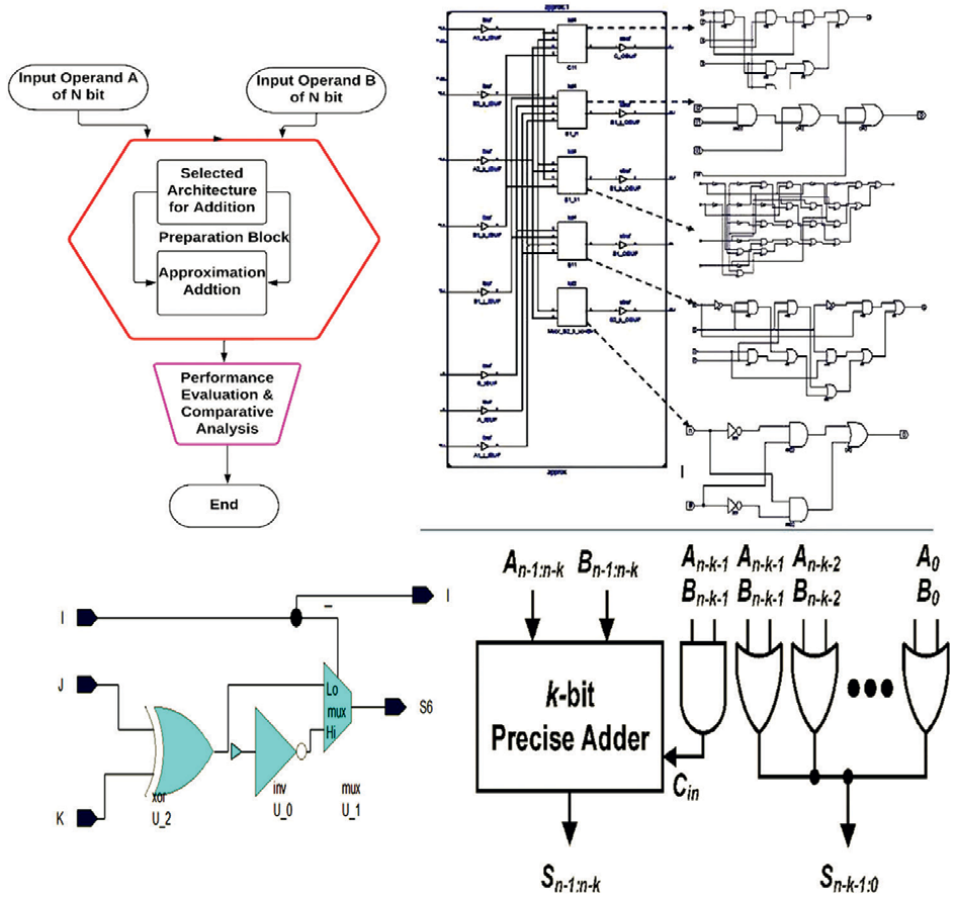

Figure 8.
(a) General flow model of approximation method and evaluation steps and (b) LUT-based implementation approximate adder.

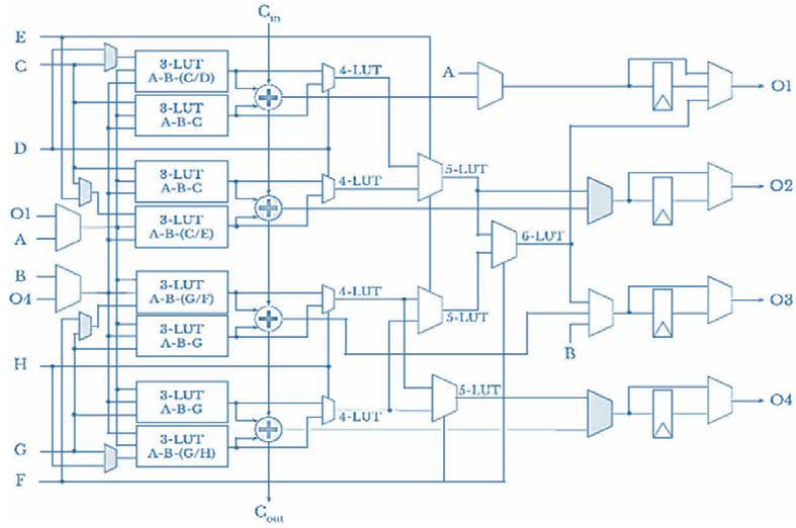

Figure 9.
PAXA-based architecture based on LUT for the 8-bit width variations.

is shown in the figure, which has two sections, approximate and accurate parts. The MSB is precise adder of K bit and LSB is imprecise adder K-L bit, where the L is length of the total sequence. The process of PAXA is simply explained by the flow chart specified [31]. Here, the N bit of A and B operand is given to the various architecture for addition. The approximation block is converting the data into K and K-L of precise and inaccurate adder. Later, the performance is analyzed based on the prime factors of edge devices such as power, area, and delay.

The PAXA is designed based on the mux logic with an XOR gate. Consider the scenario I, J, and K are the inputs provided to the PAXA. In this, "I" will act as the control switch for the multiplexer block. The XOR of J&K and XNOR of J&K inputs are the inputs multiplexer. The multiplexer is 2X1 model. Therefore, the "I" values are varied to select the desired output. These convenient representations make the FA as error redundant, and the maximum error that occurred for the FA is one out of eight test case vectors. The complete architecture of PAXA is constructed by considering the LUT of 3-bit for the inputs of A to F. **Figure 9** illustrates the PAXA architecture based on LUT designs and the corresponding output (here, represented as O1 to O4). The different approximations are performed by the LUT of 4 and 5 bits at the consecutive stages. The LUT bit width is supported for 4, 5, and 6 [32]. Gradually, the pre-trained equations are specified initially as specified in the PAXA. The output of the architecture is represented by O1 to O4. This output is generated at the end stage mux block specified in the figure.

4. Experimental results and discussions

The proposed architecture is tested in the MATLAB environment with the system specification of i7 processor, 16GB RAM with 1 TB HDD. The architecture is implemented for the 64-bit model. PAXA combined architecture of 64-bit is based on the top-level model, and it supports the efficient implementation of hardware. The experiment sections described the various approximate adders proposed for the analysis of images

or signals at the edge device. The (LPOS) -lower part and OR Summer [33], (LORA) lower part and OR Approximate [34], (HOAEM) Hardware optimized adder with error minimized [35], (ETA)-error tolerant adder [22], (HEAA)- Hardware Efficient approximate adder [36], (HEAND)-Hardware Efficient approximated adder with near error distribution [16], (MHOAEM)-Modified HOAEM, (MHEAA)-Modified HEAA PADS [37] are considered here as benchmark approximate adder circuits.

These approximated computations are compared with the PAXA, and its performance is measured by PSNR (peak signal-to-noise ratio), MAE (mean absolute error), RMSE (root mean square error value), and SSIM (structural similarity index) [38]. Also, PAXA supports for edge computing devices. It is implemented in the Vivado Platform with the ZYNQ-FPGA board. This work proposes a better PAXA for edge computing devices whose blocks of computations are formulated by adders, multipliers, and multiplier-accumulated units. This PAXA, with other models, can combine based on the edge device requirements for faster computations. These combined structures are named as Hybrid models, whose power consumption for the high-end application is minimal.

These hybrid models support both fixed and floating models with an error-tolerant level and better accuracy. To validate the performance of the PAXA, complete architecture is designed by connecting various bit widths of LUT as specified in **Figure 10**. **Figure 10** visualizes the PAXA-based overall architecture and experimental setup of DUT for edge detection and analysis in the EDGE architecture of DL. The overall architecture has two sections: (i) processing the image information and (ii) computation blocks for acceleration. The MNIST-based dataset is used for the experiment analysis of the PAXA. However, as the signals are decoded in the form of pixels of images, we considered the test jpeg image "cameraman.jpeg" for the representation. The collected images are stored in dynamic random access memory (DRAM) memory for the characteristic's representations. The I/O controller unit is responsible for converting the images into the convolution representation for deep learning models. Hence, the images from the DRAM are processed by the input-output sections of processing section 1. After this, the pixels are mapped into the matrix of 3*3. The MUL_ADD is the preprocessing block designed from the PAXA blocks. The computation speed is improved by using the multiplexer block.

The concatenation block and shift registers are present at the end section of the processing section. However, these processes are connected to the microcontroller unit (MCU). The experimentation is carried out on the ZYNQ-based FPGA, which provides sustainable solutions for speed, area, and power improvements. Section 2 is about connecting the image transformation to extract the features required for the analysis. The blocks are the I/O controller block, edge block, widow buffers, and computation blocks. The I/O block is responsible for collecting the matrix information via the first in first out (FIFO) register. The edge block is connected at the registers of the serial buffer register (SBR) and the serial buffer transform register (SBT) and these blocks are helpful in enabling the signal transmission between the edge devices and associated MCUs.

However, the speed of the computation block is independent of these register data streams. The window buffer section represents the data in the edge-segmented way to support the max pooling network of the deep learning models. Hence, the modification on the PAXA supports the image segmentations by discarding the redundant bits. **Figure 11** illustrates the outputs generated for the "camera man.jpeg" using MATLAB and categorized on PSNR and SSIM-generated images of existing models and PAXA. The bit length varies and can be incorporated with the hardware. The ZYNQ-FPGA board used for the data process bit width of 16-64 bit, which are comfortable with the

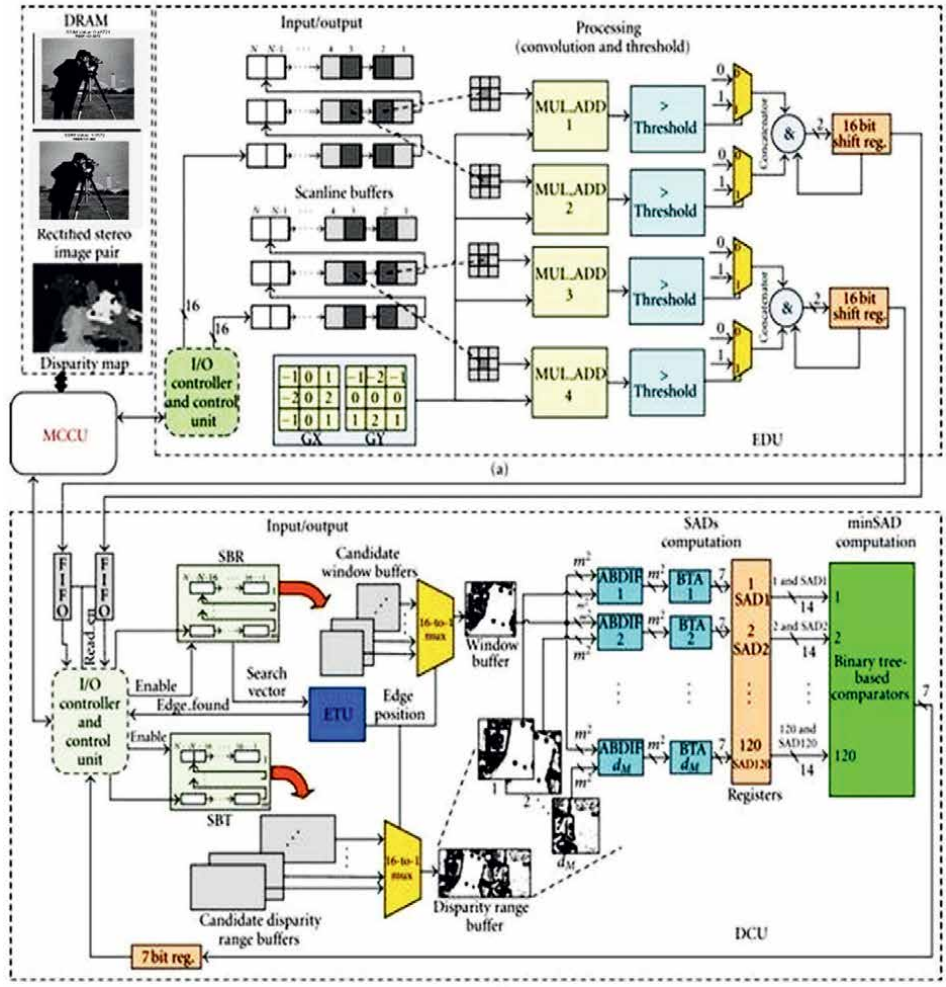

Figure 10.
PAXA-based overall architecture and experimental setup of DUT for edge detection and analysis in edge architecture of DL.

modern GPU at the cloud sections. This architectural modification ensures accelerated signal movements between the computation blocks.

The end devices observe the data variation from the buffer units. The reverse process is performed here by representing the values in the binary format, which are quickly accessed by the processors. Therefore, the usage of PAXA-based models accelerates the computation by ensuring the exact signal transfer between the blocks of shift registers and buffers. The experiment analysis is carried out on the image processing applications running at the edge device – ZYNQ. The performance metrics we consider are PSNR, MAE, SSIM, and RSME, which are closely related to represent the images. The deep learning acceleration architecture DENSENET121, LeNet, and VGA18 are considered as benchmark architecture. **Table 2** lists the comparison results on edge detection based on PAXA architectures and existing benchmark units. The conventional adder present in the DL architectures is replaced by the PAXA, and the simulations are executed to validate the performance.

Figure 11.
PSNR and SSIM generated images of existing models and PAXA.

The hardware utilization was measured from the utilization report obtained from the system reports on post-synthesis process. The hardware FPGA with PAXA is implemented under the 45 nm design technology. The results are listed in the table. The PSNR for the proposed model is 28.47, MAE is 75.17, RSME is 105.42, and the SSIM is 0.89. The PSNR of the PAXA is improved by 34.14% for LPOS, 21.95% for LORA, 16.58% for ETA, 19.81% for HEAA, 3.86% for HEAND, and 22.09% for MHEAA. The simulated results show that better PSNR is achieved by the PAXA, which enables the AI edge devices to

Types	PSNR	MAE	RMSE	SSIM
LPOS	18.75	161.54	230.72	0.45
LORA	22.22	226.77	330.98	0.53
HOAEM	23.75	197.94	263.27	0.51
ETA	22.83	170.64	215.09	0.50
HEAA	27.37	155.44	205.49	0.77
HEAND	26.41	141.23	207.98	0.52
MHOAEM	22.18	127.11	233.69	0.54
MHEAA	24.61	144.55	218.16	0.71
PAXA	28.47	75.17	105.42	0.89

Table 2.
Experimental results on edge detection based on PAXA architectures.

observe the maximum information. The MAE is reduced by 14.14% for LPOS, 20.76% for LORA, 62.58% for ETA, 27.07% for HEAA, 6.84% for HEAND, 87.93% for MHOAEM and 69.14% for EHAA. The SSIM of the PAXA is 0.89, which is much better than the other existing models, with a wide range of improvements from 39.32% to 49.43%. The hardware utilization and the device usage are based on the usage of look up table (LUT), flipflop (FF), critical path delay (CPD), area, and the clock period. The proposed model utilizes only 27 LUT and has a minimum FFT of 83. Due to this lower device utilization the CPD is reduced to 1.1 nS for the clock period of 2 nS with the area of 496.5 μm^2. Compared to the conventional adders of existing approximate methods, the proposed model's LUT is reduced by 30% for LPOS, 15% for LORA and HOAEM, 4% for ETA, 19% for HEAA, 1% for HEAD, 7% for MHOAEM, and 4% for MHEAA reducing the size of a lookup table (LUT) in very large-scale integration (VLSI) design offers several advantages that can enhance the efficiency, performance, and cost-effectiveness of integrated circuits. Firstly, a smaller LUT requires less silicon area on the chip. This reduction in area directly translates to cost savings, as smaller chip sizes reduce manufacturing costs. Additionally, it allows for more efficient use of silicon real estate, which enables designers to incorporate more functionality or features into the same chip area. Also, it supports (i) lower power consumption, (ii) improved performance, and (iii) reduced complexity of the LUT, simplifies the design process and makes the circuit easier to verify and debug.

Table 3 lists the pre-synthesis report of PXA architecture implementation for edge computing devices in IoMT. Compared to the conventional adders, the proposed model's FF usage is reduced by 33% for LPOS, 31% for LORA, 30% for HOAEM, 27% for ETA, 28% for HEAA, 19% for HEAD, 24% for MHOAEM, and 27% for MHEAA. Lowering the number of flip-flops in ICs leads to improved power efficiency and area savings, which are key considerations for optimizing performance, reducing costs, and enhancing the overall design of integrated circuits. The proposed model usage is reduced large by 24% for LPOS, 20% for LORA, 25% for HOAEM, 8% for ETA, 6% for HEAA, and 16% for HEAD. However, the PAXA has area improvement by +4% compare to MHOAEM and MHEAA-based adders. Even though the PAXA has better results on other parameters, the critical path delay in integrated circuits plays a crucial role in ensuring optimal performance and dependability of designs. This delay signifies the longest duration taken by a combinational logic path or a series of logic gates within an IC, setting the upper limit for the circuit's operational frequency and overall

Types	LUT	FFs	Clock period (nS)	Critical path delay (nS)	Area (μm^2)
LPOS	35	110	2.5	1.3	615.1
LORA	31	109	2.1	1.2	595.3
HOAEM	31	108	2.0	1.2	618.4
ETA	26	105	2.0	1.2	535.7
HEAA	32	106	1.9	1.1	526.8
HEAND	27	99	1.9	1.1	577.1
MHOAEM	29	103	1.9	1.1	475.5
MHEAA	28	105	2.0	1.2	477.4
PAXA	27	83	2.0	1.1	496.5

Table 3.
Pre-synthesis report of PXA architecture implementation for edge computing devices in IoMT.

efficiency. By minimizing this delay, integrated circuits can achieve heightened performance levels, support higher clock speeds, and exhibit enhanced timing consistency. These improvements are instrumental in bolstering the competitiveness and success of contemporary semiconductor technologies, leading to more efficient and reliable electronic systems. Path delay is reduced by 18% for LPOS; other models have a reduction of 9%. Apart from the parameters explained in the hardware utilization, additionally, we consider the improvement in CPD, area, power, area-delay product (ADP), and power delay product (PDP) after the post-synthesis process. Compared to the existing methods of approximate adder types, our proposed architecture shows promising performance in post-synthesis. The reports of device utilization after the post-synthesis have been listed in the table. The existing approximate adders are implemented in the following architecture: LPOS, LORA, HOAEM, ETA, HEAA, HEAND, MHOAEM, and MHEAA. The CPD is reduced by 16% for LPOS, 7% for LORA, 5% for HOAEM, 5% for ETA, 1% for HEAA, 1% for HEAND, 1% for MHOAEM, and 6% for MHEAA.

The CPD determines the maximum time required for a signal to propagate through the circuit. As per the tabulation, our proposed architecture reduces the latency by 16–1%, which gratefully creates an impact on IC performance to accelerate the functionality in DL applications. **Figure 12**, Illustrates the Probability value for image quality parameters based on PAXA architecture based on the results generated from **Tables 2** and **3**. The area occupancy of the proposed model after post-synthesis is reduced by 24% for LPOS, 20% for LORA, 25% for HOAEM, 8% for ETA, 6% for HEAA, and 16% for HEAND. However, the modified circuits show a lower area occupancy than the proposed work. It improves by 4% for MHOAEM and MHEAA. The remaining models consume more area than the proposed model. The power reduction is essential in accelerator circuits of VLSI to improve energy efficiency, manage thermal issues, and enable scalability and integration. By optimizing power consumption, designers can achieve higher performance, reliability, and cost-effectiveness in accelerator-based systems. Deep learning accelerators are specialized hardware designed to efficiently execute the computations involved in training and inference tasks of deep neural networks (DNNs). Given the computational intensity of deep learning algorithms, power consumption is a critical concern when designing these accelerators.

The power management ensures the improvement in scalability, thermal management, and energy-efficient circuits. The power is reduced by 27% for LPOS, 20%

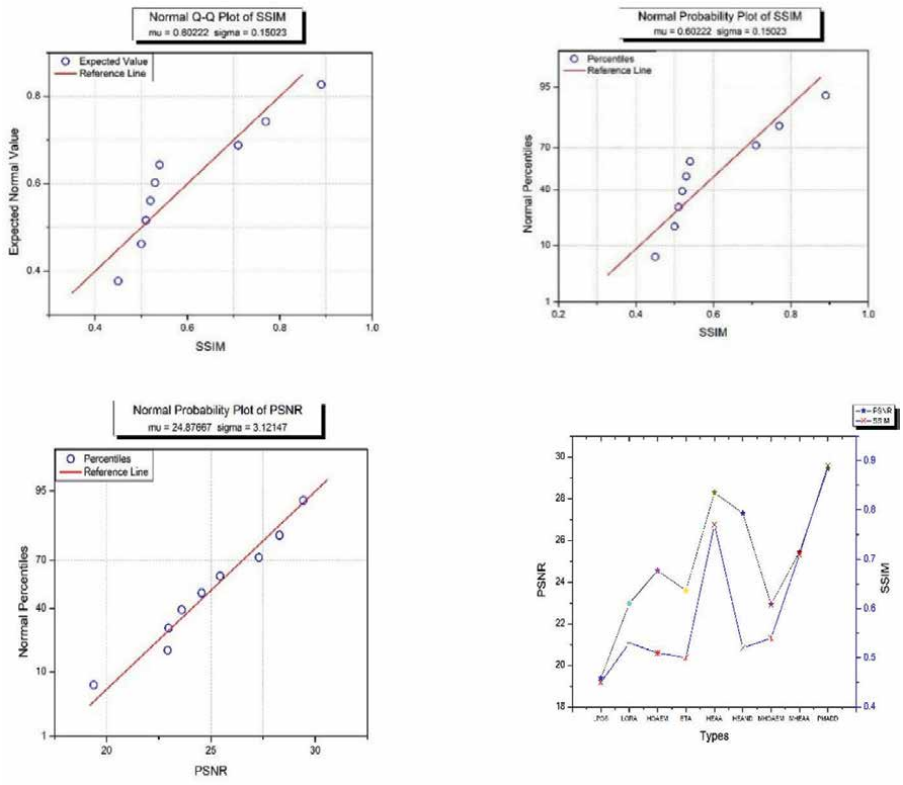

Figure 12.
Illustrations of probability value for image quality parameters based on PAXA architecture.

for LORA, 6% for HOAEM, 34% for ETA, 21% for HEAA, 45% for HEAND, 19% for MHOAEM, and 16% for MHEAA. Also, the leakage power in VLSI refers to the energy consumed by transistors when they are turned off, yet they still experience current flow because of inherent leakage mechanisms. As semiconductor technology advances with shrinking technology, nodes, and smaller transistor sizes, the issue of leakage power has escalated, posing a notable challenge in contemporary VLSI design. Therefore, the leakage power is reduced by 30% for LPOS, 64% for LORA, 52% for HOAEM, 38% for ETA, 65% for HEAA, 7% for HEAND, 10% for MHOAEM, and 34% for MHEAA. Similar to **Figure 12**, the illustrations of post-synthesis report proposed model vs. existing model based on ZYNQ-FPGA boards for deep learning – edge computing models are mentioned in **Figure 13**. The total power consumption of the architecture is closely related to the sum of static, dynamic, and leakage power. However, for the proposed architecture model is concern the total power is reduced by 27% for LPOS, 24% for LORA, 11% for HOAEM, 35% for ETA, 25% for HEAA, 51% for HEAND, 18% for MHOAEM, and 18% for MHEAA, compared to the existing models, as specified in the **Tables 3** and **4**. Especially, **Table 4** lists the post-synthesis report of PXA architecture implementation for edge computing devices in IoMT.

The other two hardware parameters are related to power, delay, and area. One parameter is the power delay product (PDP), and another is the area-delay product (ADP). The power delay product (PDP) and area-delay product (ADP) are important metrics in the design and optimization of VLSI circuits. They provide insights

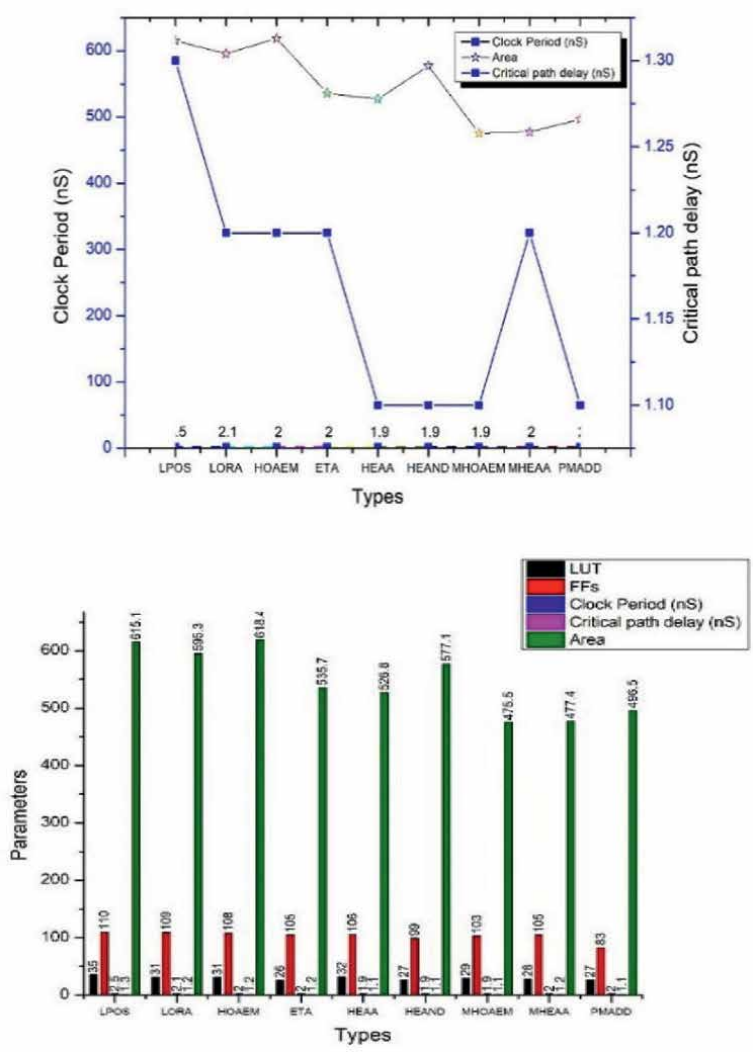

Figure 13.
Illustrations of post-synthesis reports proposed model vs. existing model based on ZYNQ-FPGA boards for deep learning – edge computing models.

into the trade-offs between power consumption, delay, and other areas, which are crucial considerations in VLSI design. The essential metrics in VLSI design that help designers optimize are energy efficiency, performance, area utilization, and scalability. By analyzing and balancing these metrics, designers can develop efficient and cost-effective VLSI solutions tailored to meet the requirements of modern electronic devices and applications. The PDP is reduced by 47% for LPOS, 33% for LORA, 16% for HOAEM, 41% for ETA, 26% for HEAA, 52% for HEAND, 20% for MHOAEM, and 25% for MHEAA, compared to the existing models as specified in **Table 4**. **Figure 14** depicts the post-synthesis reports for DL models implemented in PAXA-based FPGA boards.

Similarly, the ADP is reduced by 44% for LPOS, 28% for LORA, 31% for HOAEM, 13% for ETA, 7% for HEAA, 17% for HEAND, 3% for MHOAEM, and 2% for

Types	Critical path delay (nS)	Area (μm^2)	Power (μW)	Total power (μW)	Power delay product (f-WS)	Area-delay product (f-m^2S)
LPOS	1.30	609.62	121.92	135.66	158.72	713.26
LORA	1.20	589.94	115.63	132.98	143.61	637.14
HOAEM	1.18	612.92	102.15	118.18	125.26	649.69
ETA	1.18	530.87	129.16	143.71	152.34	562.73
HEAA	1.13	522.13	116.03	133.43	136.10	532.58
HEAND	1.13	571.91	139.39	161.26	164.48	583.34
MHOAEM	1.13	471.28	114.95	126.60	129.13	480.71
MHEAA	1.19	473.17	111.88	126.06	134.89	506.29
PAXA	1.12	492.04	96.28	106.84	107.90	496.96

Table 4.
Post-synthesis report of PXA architecture implementation for edge computing devices in IoMT.

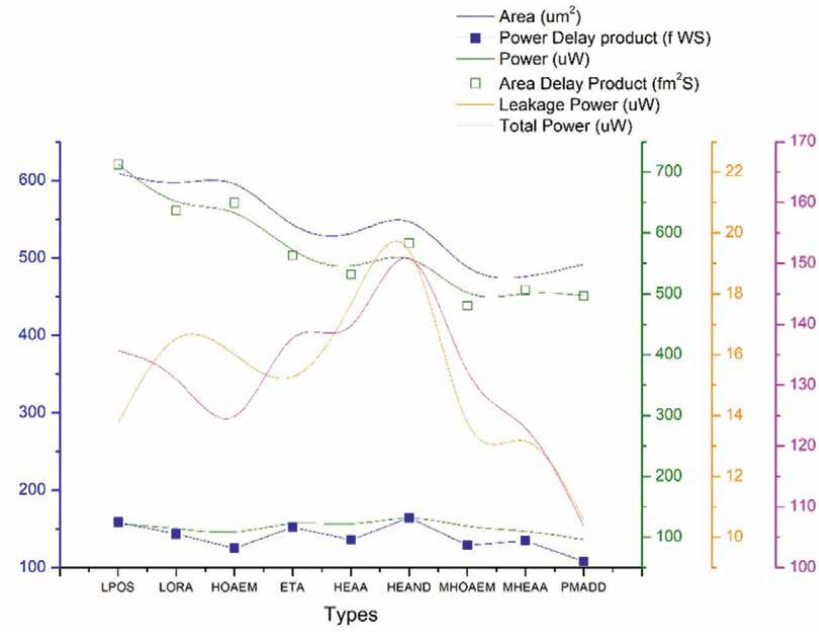

Figure 14.
Illustration of post-synthesis reports for DL models implements in PAXA-based FPGA boards.

MHEAA, compared to the existing models as specified in the table. The corresponding visualizations are illustrated at the appropriate places. After several trials and investigations on the proposed model, we observed that this model is highly suitable for image processing deep learning models such as CNN, RNN, and DBN. The suggested accumulator arithmetic block offers a promising solution for AI applications that can tolerate errors, such as deep neural networks (DNNs) used in tasks like image classification, object detection, and various image processing tasks.

4.1 Opportunity of IoMT beyond healthcare's

The Internet of Medical Things (IoMT) can be generalized or adapted to contexts beyond healthcare and is crucial for innovation and cross-disciplinary application. The IoMT technologies can be extended to other domains: (i) Data Security and Privacy: The robust data security and privacy measures developed for IoMT can be adapted to other fields dealing with sensitive information, such as finance, government, and personal communication. Techniques like encryption, blockchain, and secure authentication protocols have broad applicability. (ii) Remote Monitoring and Control: The remote monitoring and control systems used in healthcare can be repurposed for various industrial applications. (iii) Predictive Analytics: The predictive analytics algorithms employed in IoMT to forecast patient outcomes and suggest treatment plans can be applied to other domains, like supply chain management for predicting demand, optimizing logistics, and reducing waste. (iv) Wearable Technology: Wearable devices designed for health monitoring can be adapted for fitness tracking, sports performance analysis, and workplace safety. These devices can track vital signs, movement patterns, and environmental factors to improve performance and prevent injuries. (vi) Smart Infrastructure: The infrastructure supporting IoMT, such as interconnected devices and communication networks, can be applied to smart cities and environmental monitoring. By deploying sensors and IoT devices, cities can optimize resource utilization, manage traffic flow, and monitor air and water quality. The user-centric design principles and ethical implications surrounding the deployment of such technologies in medical settings.

4.2 User-centric design principles of IoMT

The user constraints principles of IoMT are discussed along with the ethical implications as follows. The user-centric design principles and ethical implications surrounding the deployment of such technologies in medical settings. User-centric design principles: (i) Accessibility: ensure that IoMT devices and interfaces are accessible to all users, including those with disabilities. This involves designing user interfaces with clear navigation, large fonts, and compatibility with assistive technologies. (ii) Usability: prioritize simplicity and ease of use to accommodate users with varying levels of technical expertise. Intuitive interfaces, clear instructions, and minimalistic design can enhance usability and user satisfaction. (iii) Privacy and consent: implement robust privacy controls and obtain informed consent from users regarding data collection, storage, and sharing. Transparent privacy policies and granular consent options empower users to make informed decisions about their data, data integrity, and interoperability. As far as the ethical implications are concerned, data privacy and informal contests are framed to adhere to data protection regularity by ensuring the safety of services and autonomous decisions about healthcare decisions. Hence, it supports the bias and fairness with the transparency and accountability of the IoMT technologies including decision-making algorithms and ethical implications by incorporating user-centric design principles and addressing ethical considerations, healthcare providers can deploy IoMT technologies responsibly, promoting patient-centric care while safeguarding privacy and autonomy.

5. Conclusions

The IoMT field requires faster computation and reliable performance. This is achieved by the ECB design. In this research work, we propose the ECB, whose

architecture is modified based on the approximate computations. This proposed approximate adder (PAXA) ensures better performance in edge computing devices, especially deep learning architectures. As our results show the promising improvements in prime factors of chips, it is highly suitable for accelerating the performance of the edge devices. The experiment investigation is carried out by setting up the deep believe networks environment in the ZYNQ-based FPGA. The results of edge detection and image processing under these models of approximated adder architecture achieve the PSNR of 28.47, MAE of 75.17, RMSE of 105.42, and SSIM of 0.89. These results are marginally high and show better improvement in all the aspects of edge device requirements. Compared to the existing models and benchmark circuits, our PAXA shows enhanced performance for ECB-based accelerated beep believe network. Similarly, during the implementation process, the models show reliable performance such as CPD of 1.12 nS, area of 492.04 μm^2, power of 106.84 μW, ADP of 496.96 f-m2s, and PDP of 107.9 f-WS. Therefore, these models are highly recommended for improved accelerated computation units of edge devices. Also, it supports low latency cycle and low-power consumption required for high-end power transmissions, especially in autonomous vehicles, robotic surgery, diagnosis, and medicine distributions. By implementing the proposed adder (PAXA) at the high end – edge computing gets around the dependencies by locating data that is closer to the possibility, which speeds up applications and improves their availability.

Acknowledgements

This work was proposed by Dinesh Kumar J R, and all the authors (Balaji V R, Joby Titus T, Karthi S P, and Visvesvaran C) have made equal contributions to this work and acknowledge SKCET for providing opportunity to complete this work.

Conflict of interest

The authors declare no conflict of interest in contest to this proposed work and the research outcomes.

Author details

Dinesh Kumar Jayaraman Rajendiran*, Balaji Venkatesalu Ramasamy, Joby Titus T, Karthi Samiyampalayam Palanisamy and Visvesvaran Chandramohan
Sri Krishna College of Engineering and Technology, Coimbatore, India

*Address all correspondence to: dineshkumarjr@skcet.ac.in

References

[1] Xue H, Huang B, Qin M, Zhou H, Yang H. Edge computing for internet of things: A survey. In: International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics). Rhodes, Greece: IEEE; 2020. pp. 755-760. DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00130

[2] Wang X, Ren X, Qiu C, et al. Synergy of edge intelligence and blockchain: A comprehensive survey. TechRxiv. Springer; 2021;**2**(10):102-143

[3] Yu W et al. A survey on the edge computing for the internet of things. IEEE Access. 2018;**6**:6900-6919. DOI: 10.1109/ACCESS.2017.2778504

[4] Khan WZ, Ahmed E, Hakak S, Yaqoob I, Ahmed A. Edge computing: A survey. Future Generation Computer Systems. 2019;**97**:219-235. ISSN 0167-739X. DOI: 10.1016/j.future.2019.02.050

[5] Jain P, Huda S, Maas M, Gonzalez JE, Stoical I, Mirhoseini A. Learning to design accurate deep learning accelerators with inaccurate multipliers. In: 2022 Design, Automation and Test in Europe Conference and Exhibition (DATE), Antwerp, Belgium. Springer. 2022. pp. 184-189. DOI: 10.23919/DATE54114.2022.9774607

[6] Wang X, Wang L, Chu Z, Xia Y. Design and evaluation of approximate adders. In: 2020 IEEE 14th International Conference on Anti-counterfeiting, Security, and Identification (ASID), Xiamen, China. IEEE; 2020;**2**:201-204. DOI: 10.1109/ASID50160.2020.9271703

[7] Nithyashree RV, Afreen S, Tantry S. Analysis of various approximate adders in ripple carry adder design. In: 2023 Fourth IEEE Global Conference for Advancement in Technology (GCAT), Bangalore, India. IEEE. 2023;**2**:1-5. DOI: 10.1109/GCAT59970.2023.10353293

[8] Smith A et al. Efficient approximate adders for FPGA-based edge computing. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2020;**9**(2):100510-100521

[9] Chen B et al. Stochastic computing based approximate adders for low power edge devices. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2019;**1**:1-5. DOI: 10.1109/ISCAS.2019.8702248

[10] Sarkar MR, Yi CY. An In-Memory Computing Architecture Utilizing Energy-Efficient VGSOT MRAM Device. IEEE Transactions on Circuits and Systems II: Express Briefs. July 2021;71(7):3258-3262. DOI: 10.1109/TCSII.2024.3359993

[11] Zhang L, Wang Y. Approximate adders for high-throughput edge computing applications. IEEE Transactions on Emerging Topics in Computing. 2018;**1**(10):2313-2325

[12] Liu X, Li Z. Error-resilient adder designs for ultra-low power edge devices. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. 2022;**1**(1):913-916

[13] Tripathi D, Wairya S. An energy dissipation and cost optimization of QCA ripple carry adder. In: 2021 Eighth International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India: IEEE; 2021;**4**:760-765. DOI: 10.1109/SPIN52536.2021.9566068

[14] Daniel Raj A, Saravana Kumar R, Deb S, Vignesh Roshan M, Sugirdan V, Soundar S. Design and analysis of high-performance carry skip adder using various full adders. In: 2021 Smart Technologies, Communication and Robotics (STCR). Sathyamangalam, India: IEEE; 2021;**3**:1-5. DOI: 10.1109/STCR51658.2021.9588863

[15] Rosa M et al. AxPPA: Approximate parallel prefix adders. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2023;**31**(01):17-28. DOI: 10.1109/TVLSI.2022.3218021

[16] Mohsen V, Pavel L, Ali NB. Design and implementation of novel efficient full adder/subtractor circuits based on quantum-dot cellular automata technology. Applied Sciences. 2021;**11**:8717

[17] Amira G, Ihsen A, Khaled K, Mouna B, Tarek F, Mohamed A, et al. Defensive approximation: Securing CNNs using approximate computing. In: Proceedings of the 26th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'21). ACM, New York, NY: IEEE; 2021;**21**(7):7990-8003

[18] Cao K, Liu Y, Meng G, Sun Q. An overview on edge computing research. IEEE Transactions. IEEE; 2020;**8**:85714-85728. DOI: 10.1109/ACCESS.2020.2991734

[19] Manohar PS, Rohan B, Ramana PVS, Jamal K, Kumar MOVP, Reddy BV. Implementation of carry look ahead adder with 2-bit approximate adder. In: 2023 Second International Conference on Applied Artificial Intelligence and Computing (ICAAIC). Salem, India: IEEE; 2023;**1**(1):1543-1547. DOI: 10.1109/ICAAIC56838.2023.10140683

[20] Trivedi V, Lalwani K, Raut G, et al. Hybrid adder: A viable solution for efficient design of MAC in DNNs. Circuits, Systems, and Signal Processing. 2023;**42**:7596-7614. DOI: 10.1007/s00034-023-02469-1

[21] Mishra V, Mittal S, Hassan N, Singhal R, Chatterjee U. VADF: Versatile approximate data formats for energy-efficient computing. ACM Transactions on Embedded Computing Systems. 2023;**22**(5s):1-21. Online publication date: 31-Oct-2023

[22] Seo H, Yang YS, Kim Y. Design and analysis of an approximate adder with hybrid error reduction. Electronics. 2020;**9**:471. DOI: 10.3390/electronics9030471

[23] Sergi A, Albert C-A, Eduard A, Josep T. WiSync: An architecture for fast synchronization through on-chip wireless communication. ACM SIGARCH Computer Architecture News. 2016;**44**(2):3-17

[24] Jair C, Farid G-L, Lisbeth R-M, Asdrubal L. A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing. 2020;**408**:189-215

[25] Bastien D, Marcello T, Arnaud V, Patrick G. Reducing overprovision of triple modular redundancy owing to approximate computing. In: Proceedings of the 27th International Symposium on On-Line Testing and Robust System Design (IOLTS'21). Los Alamitos, CA: IEEE; 2021;**21**(7):2045-2067

[26] Rasoul FS, Pierre A, Kia B. Approximate constant-coefficient multiplication using hybrid binary-unary computing for FPGAs. ACM Transactions on Reconfigurable Technology and Systems. 2021;**15**(3):1-25

[27] Sabireen H, Neelanarayanan V. A review on fog computing: Architecture,

fog with IoT, algorithms and research challenges. ICT Express. 2021;**7**(2):162-176. ISSN 2405-9595. DOI: 10.1016/j.icte.2021.05.004

[28] Michael J, Marcelo B, Guilherme M, Geraldo O, Arthur L, da Bruno S, et al. Data clustering for efficient approximate computing. Design Automation for Embedded Systems. 2020;**24**(1):3-22

[29] Bapi K, Kumar GP, Kumar BS, Mohendra R, Arindam B. ADIC: Anomaly detection integrated circuit in 65-nm CMOS utilizing approximate computing. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2020;**28**(12):2518-2529

[30] Masoomeh M, Shahriar SH. Energy efficient 3D network-on-chip based on approximate communication. Computer Networks. 2022;**203**:108652

[31] Luca NG, Pedro K, Marcos L, Ben J. Lightweight dual modular redundancy through approximate computing. In: Proceedings of the XI Brazilian Symposium on Computing Systems Engineering (SBESC'21). Los Alamitos, CA: IEEE; 2021. pp. 1-8

[32] Burks AW, Goldstine HH, Neumann JV. Preliminary discussion of the logical design of an electronic computing instrument. In: The Origins of Digital Computers. Berlin, Heidelberg: Springer; 1982. pp. 399-413

[33] Mrazek V, Sarwar SS, Sekanina L, Vasicek Z, Roy K. Design of power-efficient approximate multipliers for approximate artificial neural networks. In: 2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), Austin, TX, USA. ACM; 2016;**1**(4):1-7

[34] Jiang H, Angizi S, Fan D, Han J, Liu L. Non-volatile approximate arithmetic circuits using scalable hybrid spin-CMOS majority gates. IEEE Transactions on Circuits and Systems I: Regular Papers. 2021;**68**(3):1217-1230

[35] Hanif MA, Hafiz R, Hasan O, Shafique M. PEMACx: A probabilistic error analysis methodology for adders with cascaded approximate units. In: 2020 57th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA, USA. IEEE; 2020;**1**(4):1-6

[36] Bhargav A, Huynh P. Design and analysis of low-power and high speed approximate adders using CNFETs. Sensors. 2021;**21**(24):8203. DOI: 10.3390/s21248203

[37] Erfan FS, Mohammad RR. Power-efficient, high-PSNR approximate full adder applied in error-resilient computations based on CNTFETs. In: Proceedings of the 2020 20th International Symposium on Computer Architecture and Digital Systems (CADS). Rasht, Iran: IEEE; 19-20 August 2020;**2**(3):112-125

[38] Hamdan S, Ayyash M, Almajali S. Edge-computing architectures for internet of things applications: A survey. Sensors. 2020;**20**:6441. DOI: 10.3390/s20226441

Section 2

Application Scenarios

Chapter 4

Safety Assurance in IoT-Based Smart Homes

Mouiad Al-Wahah and Auhood Al-Hossenat

Abstract

A smart home's safety is a very urgent question due to several causes. This chapter analyzes current directions of smart house system safety technologies in use nowadays. Current studies are dedicated to the integration of Internet of Things (IoT) into smart home systems; critical situations that may arise; and specifications of sensors in the smart home system. The huge number of connected devices and the capacity embedded within these devices to direct demand resources make deliberate attacks on them and/or inadvertent downfall events such as abrupt bad interactions between connected devices, mechanical failure of devices, and unsuccessful communication may lead to IoT-based systems entering unreliable and threatening physical states. We review current trends in security-enabled safety monitoring frameworks for IoT-based smart homes. We demonstrate the use of various techniques in utilizing system analysis during design to develop a monitoring model that can be executed, providing run-time safety assurance for a system. This is achieved through collecting and analysis of operational data and evidence to assess the safety status of the system. Subsequently, appropriate actions are taken, and the safety status is communicated securely to system users, along with recommended actions to reduce the risk of the system entering an unsafe state.

Keywords: safety assurance, smart home, artificial intelligence, internet of things, simulation, modeling

1. Introduction

The Internet of Things (IoT) has occupied several aspects of our daily lives; specifically, it turns our place of right-mindedness and convenience into a pool of interacting smart devices that allow such comfort, adaptability, automation and stability. The rise of intelligent technologies in households offers a range of services and functionalities for people's lives. Although a smart home (SH) offers significant opportunities for increased comfort and risk management, it presents new safety hazards and it also changes the nature of the already existing risks [1].

Modern smart home relies heavily on a mixed of hardware and software-intensive systems. These intensive systems include safety-critical systems such as urgent-care medical equipment, gas leakage detection, fire alarming system, and so on. In many of these systems, an abrupt problem in the software or hardware can lead to

hazardous failures with the potential for loss of life. Since safety is a key requirement for the SHs, failures in safety-critical systems carry the potential for serious and disastrous impacts, including posing risks to human life or even causing loss of life [2]. Because of this delicate relationship between SHs and hazards originating from techniques these SHs employ, safety assurance and risk analysis techniques have become of paramount necessity in designing and operating of SHs.

Safety assurance is the planned and systematic process to ensure adequate confidence in the safety of a product, a service, or a functional system [3]. It represents the way to create a safety argument and to prove that this argument always indicates safe conditions [4]. Several definitions of safety have been introduced in the literature, for example, the Cambridge dictionary defines it as "a state in which or a place where you are safe and not in danger or at risk" [5]. However, the most appropriate one is given by Dezfuli et al. [4] when they define the safety as "freedom from conditions that may result in death, injury, occupational illness, equipment and property loss, or harm to the environment".

The study's assessment method is based on the analysis of works published in previous authentic, peer-reviewed, and famous scientific conferences and journals that are indexed in relevant scientific databases in addition to survey studies for distinguished related works.

The study provides a larger scope in the field of IoT safety than the previous studies; hence, it can be productively used by future researchers in the smart home safety assurance and give handy and deeper comprehension and guidance for the IoT-based smart home topic's researches and professionals. The study's main contribution is providing an updated literature revision on Safety Assurance (SA) approaches for Smart Home (SH) as a Cyber-Physical System (CPS), with a focus on the two main frameworks in this concern: design-time SA and run-time safety assurance. To reach this goal, the study is organized as follows: Section 2 explains the IoT architecture as introductory information for the reader. Section 3 provides the works related to this study. Section 4 presents challenges of Smart Home Safety Assurance (SHSA). Section 5 is reserved for SA problem. In Section 6, we conclude our study.

2. IoT architecture

Before discussing safety assurance in IoT-based SH, the existing layers of the IoT architecture will be briefly highlighted. The term IoT literally stands for "Internet of Things" and it refers to any combination of devices (things) that are connected to the Internet [6]. However, numerous intelligent devices rely on proxies like hubs or local connections via Bluetooth or Zigbee wireless instead of directly connecting to the internet, Radio Frequency Identification (RFID), or Wi-Fi. Wi-Fi provides enhanced ways to connect objects together and enable communication not only among themselves but also with the internet. For the sake of this study, the term IoT refers to any group of functions that includes at least two physical components that can be connected to over any kind of network. The Internet of Things utilizes specific protocols with sensing devices to facilitate communication for smart recognition, positioning, tracking, monitoring, control, operation, and management [7]. There are various IoT architectures available for IoT devices. In this study, we are going to depend on the 4 layers paradigm since it is simple and comprehensive.

2.1 Perception layer

The IoT architecture's perception layer consists of a variety of devices that focus on sensing the environment and activating physical processes. Various devices and technologies that receive information from the surroundings are included in the perception layer. Pressure sensors, smoke sensors, vibration sensors, and RFID sensors are some of the devices and technologies available [5] to precept physical parameters, such as object properties, biometrics, and physiological or environmental conditions. Moreover, this layer includes actuators that work according to commands coming from processing layer. These devices are anticipated to possess a high level of dependability, user-friendliness, increased clarity, heightened responsiveness, intelligent detection, minimal energy usage, and other features [6].

2.2 Network layer

The network layer, the second in the IoT architecture, ensures the dependable transfer of data from the perception layer to the computational unit for processing sensing data [7, 8]. The network layer transports data through interfaces and gateways using communication technologies and protocols [9]. This level of the IoT structure establishes guidelines for collecting data. The network layer combines devices like hubs, switches, gateways, along with technologies like Bluetooth, Wi-Fi, and Long-Term Evolution (LTE) [10].

2.3 Processing layer

The data-processing layer in the IoT system is responsible for processing events, enabling smooth software communication for storing and handling IoT data [11–14]. The processing layer serves as a link connecting the application and network layers, carrying out tasks such as data accumulation, abstraction, and analysis [15]. Data are processed through cloud computing and multiparty computation, enabling both bulk data processing and intelligent handling. The layer uses machine learning, deep-learning algorithms, and data processing elements to analyze the data from the perception layer, creating new insights and sometimes predicting hazards and issuing warnings.

2.4 Application layer

The top layer of the IoT architecture is the application layer, which provides personalized services based on the end-users' specific needs [16]. The application layer serves as a bridge between external applications. The layer acts as the main connection between the users and the applications. It processes the data received from the network layer to provide the services required by the customer. The layer decodes patterns found in IoT data, then translates them into easy-to-understand summarized patterns displayed in graphs, tables, and pictorial formats for users (**Figure 1**) [17, 18].

3. Related works

Security gets the lion's share of research but safety is barely left with crumbs. Several studies have been conducted on the subject of smart home security/safety

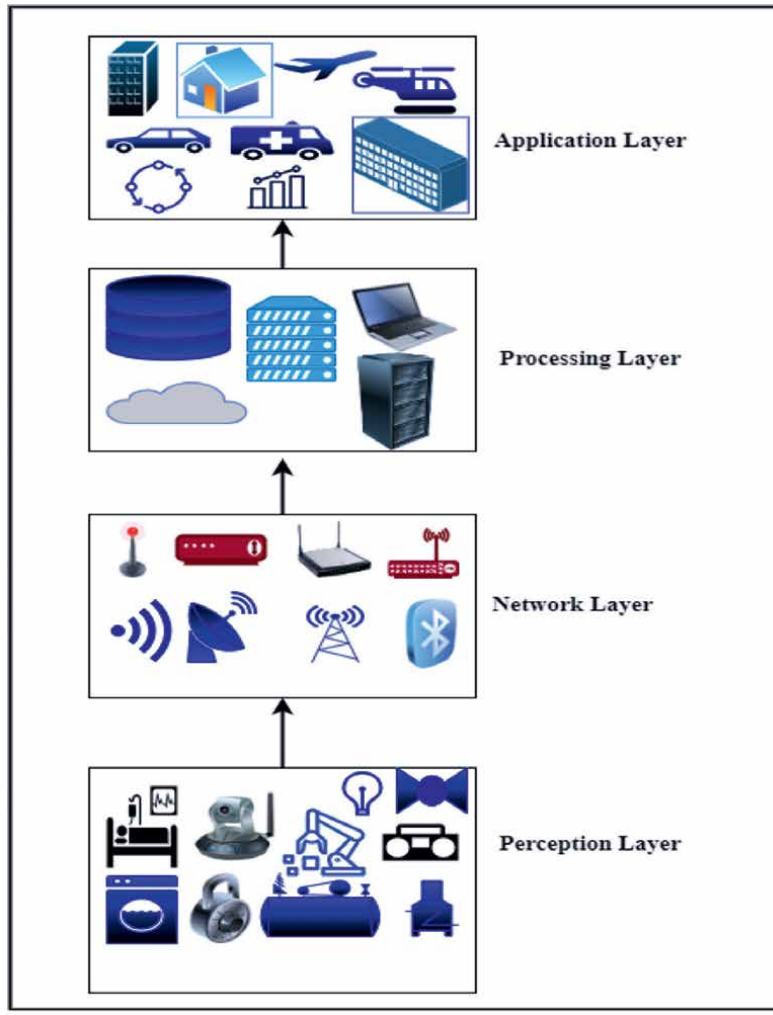

Figure 1.
IoT architecture.

assurance; however, most of them focus on security and privacy issues and neglecting, almost totally, safety concerns within SH.

Safety assurance and risk analysis have been previously discussed, with a primary emphasis on System Theoretic Process Analysis (STPA), a well-known dynamic method. A more detailed explanation of the method was provided, as well as its extensions aiming to enhance hazard analysis in intricate systems [19]. Our research, seen from a different perspective, examines primary static and dynamic safety assurance methods, emphasizing their benefits and limitations.

Another study offers a comprehensive overview emphasizing issues of reliability present in the functioning of IoT layers. It explains models trying to depict the criteria for system failure in a logical and organized way [20], but it does not address ensuring the safety of smart homes. In contrast, our research focuses primarily on investigating safety concerns associated with IoT-enabled smart homes.

A detailed examination of the security of the smart home ecosystem is conducted in a study [21]. The authors have looked into various cyber-attacks and threats that could disrupt the proper operation of various devices and services in smart homes, ultimately impacting safety. However, safety concerns were not considered in that study.

Challenges in smart home technologies enlisted by researchers were 13 [22], however, safety is not discussed. The analysis given in [22] focuses on the technical elements of these obstacles.

A closer survey to our study, with some differences, is presented by Abdulhamid et al. [23]. The researchers give a comprehensive explanation of the interwoven nature of security and safety in IoT system. The study clarifies that safety and security share four types of interplays: conditional dependency, mutual reinforcement, antagonistic relationship, and independent relationship [24, 25]. However, our study focuses on the fourth type of these interplays through explaining the major approaches used for safety assurance in isolation of the security factor.

Program analysis that is used to confirm security and safety aspects of IoT applications have been offered and suggested a variety of categorization and classification characteristics to improve comprehension of safety/security research areas. Additionally, obstacles examined are addressed in the research and the possible strategies that could be implemented to safeguard the security and safety of IoT systems [26]. Our research attempts, anyway, tries to split safety assurance in smart home from security issues and to draw a slim border between runtime and deign-time safety assurance.

Breaches in IoT security happen when threats take advantage of weaknesses in hardware or software. However, IoT safety breaches typically occur due to computer malfunctions caused by hardware or software issues from risks [27], yet basic safety concepts do not offer a comprehensive examination of SA methods. Our research, though, provides a comprehensive explanation of SHSA methods.

4. Smart home safety assurance challenges

4.1 Huge data

The tremendous amount of data generated and transmitted between connected devices has affected researchers interested in dynamic safety assurance of the SH's strategies, as it is difficult for ML approaches to deal with such big data in a timely manner. It is estimated that a full 90% of all the data in the world has been generated over the last 2 years [28].

4.2 Complexity

Complexity expresses the growing unpredictability of the system's behavior, which may jeopardize its safe and reliable operation [29]. Sharing data and connecting devices in smart homes requires using the Internet of Things as a backbone for communication. The Internet of Things comprises various levels and layers of software/hardware along with standard protocols. Due to the significant rise in shared data and connected devices, the complexity of used software/hardware and

standard protocols will also increase [30]. This leads to increased unpredictability and hence more hazards.

4.3 Cybersecurity

Cybersecurity embraces both security and privacy. There is a strong relationship between safety and cybersecurity in Cyber-Physical Systems (CPSs), including smart homes. This relationship can be characterized as a mutual dependent coexistent relationship. This is due the fact that cyber-attacks can benefit from shortfalls in the protection systems, protocols, or human careless disregard for consequences and directly influence the integrity or availability of the data and control systems [31, 32]. For example, a thief can steal a house if he can hack the cameras and control them. There are already tremendous works that meant to deal with cybersecurity challenges in IoT [33–40], and these studies have detailed the cybersecurity issues and their alleviation methods. Hence, in this study, we are not covering this subject any further.

4.4 Safety

Irrespective of safety issues related to cybersecurity, safety issues in SHs are numerous and they accommodate a wide spectrum of hardware and software problems. Smart homes include several smart things, and these things can create unsafe conditions and increase the hazard of harm to persons and property if their collective operation goes into unpredictable situation. The halt in a system's capability to carry out a necessary task or its failure to operate within set boundaries results in harm or damage [27]. For example, consider this scenario:

$$\begin{aligned} &Leaked-gas\ detector\ is\ ON + User\ is\ at\ Home + It\ is\ Morning \\ &\quad + User\ want\ to\ drink\ coffee + Speaker\ On + Stove\ is\ On \rightarrow Safe\ status \end{aligned} \tag{1}$$

$$\begin{aligned} &Leaked-gas\ detector\ is\ ON + Leaked-gas\ detector\ is\ idle + User\ is\ at\ Home \\ &\quad + It\ is\ Morning + User\ want\ to\ drink\ coffee + Speaker\ On \\ &\quad + Stove\ is\ On \rightarrow Hazardous\ status \end{aligned} \tag{2}$$

An inadequate safety legalization and standard in IoT systems also adds difficulties to this safety process. For example, unknown life times; given that Things may outlive their maintainers and the software used to control them, we need to consider how to manage Things that last longer than expected.

4.5 Human factors

These originate from human errors, hardware/software design shortcomings, confusion, and misunderstanding of the system during all system lifecycle stages. For example, inadequate requirement specifications, design mistakes, coding errors, operation pitfalls, hardware/software misconfiguration, and incomplete/erroneous software updates, to name just a few. These factors affect system behavior and may lead it to enter erroneous status.

5. IoT smart home safety assurance problem

5.1 IoT smart home safety assurance problem setup

As any automatic system, a smart home relies on learning-enabled components (LECs) [41]. These components are supplied with sensors and are typically controlled using event-driven software applications. The software applications receive their inputs as sensed data from these sensors, extended triggers, from the Cloud (or internet), user inputs, or any combination of two or more of these data feeders. Consequently, the controller (mostly resided in a central hub or any processing-capable device) software application issues a command to one or more actuators to provide different forms of automation, and this must be accompanied by checking the safety status of the whole smart home system.

At any moment, the safety status of the smart home may change from a safe status to a vulnerable status due to any intentional targeting (privacy and security attacks)), or due to an unintentional failure like design errors, configuration mistakes, updating failures, operational errors, or communication malfunctioning (most safety problems in Io-T smart homes are unintentional). After entering the vulnerable status, the system moves forward to a hazardous status (if there are no curbs/controls to alleviate the hazard and return the system to the safe status, or at least, to the vulnerable status), in which the system is ready to slip into consequence status. Once the system is in the consequence status, there will be (almost) only a little to do about safety assurance because the damage has already occurred. The moment of entering the vulnerable status represents a decisive point in smart homes since it is at this stage the system should be able to recognize the causes of this vulnerability and be prepared to deal with several situations to get the system back to the safe status. **Figure 2** illustrates the IoT-based SH safety status transitions.

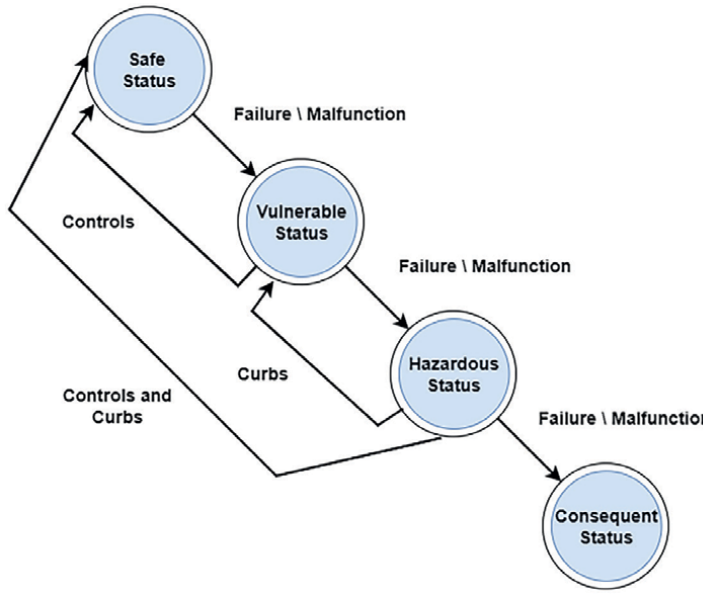

Figure 2.
IoT-based SH safety status transitions.

The fundamental concepts that are related to the smart home safety assurance are:

1. Failure is defined as the ways in which something might went wrong.

2. Vulnerability it is the exposure of the system to the possibility of being under any kind of hazard.

3. A hazard is a condition that could lead to an undesirable situation for the system. For example, undetected leaked gas could lead to an explosion.

4. Consequence is an undesirable and unexpected situation for the system [42]. For example, an explosion resulted from SH device failure.

5. Curbs are components required to prevent the hazards from leading the system to a consequence. For example, a gas clogging valve can prevent an explosion.

6. Risk is a combination of the probability (or frequency) of hazard occurrence and the severity (or consequence) [42].

5.2 Smart home safety assurance frameworks

A smart home is a complex system that consists of hardware and software components. All hazards in SH are caused by hardware component because software, as stated by Leveson [43] and Abdulkhaleq et al. [44] "software by itself is not hazardous and cannot directly cause damage to human life or environment; it can only contribute to hazards in a system context." Software has the ability to generate dangerous system conditions either by manipulating the system's fault controls or by misleading the system's human operators during their decision-making process. Safety assurance frameworks can be classified into two categories: static and dynamic. Each of these frameworks has its own advantages and disadvantages. These methods rely on creating views by analyzing thorough models of the systems' static and dynamic behaviors using techniques from available modeling languages features. Most present frameworks have been developed using the unified modeling language (UML) or system modeling language (SML) [23, 45, 46].

5.2.1 Static smart home safety assurance frameworks

This type of safety assurance techniques is achieved during the design phase of the intended SH system's System Development Life Cycle (SDLC). The reader is referred to Sommerville [47] for detailed information about software design phases. They are also called static or reliability-based safety assurance frameworks because their effect is lost once the system enters its real operation. In these methods, the system subjected to a series of testing-verification-validation processes to ensure that the final product will work in an accepted level of operational safety. Main drawbacks of this type of SA frameworks are:

1. They use qualitative data that might be biased,

2. The injected modeling/simulation/testing data may not be enough for capturing real system behavior,

3. Many design-time assumptions mismatch with real run-time environment uncertainties,

4. Most of these approaches do not provide mitigation/alleviation risk strategies.

The main advantages of these approaches can be summed up by:

1. They are closer to software design philosophies,

2. Easy to grasp, comprehend and implement,

3. Can serve as a vital preemptive defense against safety violations.

4. As it the case always, they are the input for runtime safety assurance approaches.

Ericson [48] presents a system safety technique used only for identifying expected hazards at the early design level when there is not enough detailed design information available; he named it Preliminary Hazard Analysis (PHA).

Currently, the Fault-Tree Analysis (FTA) [49] method is one of the most commonly utilized approaches for conducting safety analysis. The goal of an FTA is to identify and follow the impact of a system-level danger on separate failures of specific system parts and sub-parts. The approach employed by [50] creates a thorough security hierarchy following the overall structure of a smart home. Later on, the technique is assessed in a scenario involving successful breaches on a lightbulb network operating via the ZigBee protocol.

Saeed et al. [51] presents static approach for IoT-based intelligent home fire prevention system using multiple sensors. They use simulation technique to simulate fire in a smart home using the Fire Dynamic Simulator (FDS).

Failure Modes, Effects, and Criticality Analysis (FMECA) is developed by NASA as an extended version for Failure Modes and Effects Analysis (FMEA). It is a design-time safety assurance approach that assigns a criticality ranking for each failure. Several recent researchers use this approach for safety analysis tasks. For example, it has been used by [52, 53]; the earlier use it for risk assessment of medical devices while the later use it after amalgamating dynamic wavelet neural network with it for prognostic devices fault prediction.

In this category of SHSA, the most recent approach is given by [54] when the researchers try to map the problem of safety assurance for the IoT system into a model checking problem. They design a framework to work with Samsung SmartThings platform and named it IOTSAN (for IOT sanitizer) that catches, as they claim, IoT safety violations. No mitigation measures are given in IOTSAN system, and it is only a diagnostic tool.

In the same manner, authors in [55] have turned the safety assurance problem in SH into a model checking problem, and they call their framework safe Internet of Things (SIFT). SIFT receives user's program of IoT app with a series of event-based rules from multiple users. By aggregating rules together, complex system behavior analysis can be reached using backward chaining strategy. If any safety-related conflict is detected, suggestions for changing the user rules are issued.

SOTERIA is a static analysis system for IoT apps that performs model checking. It automatically generates a state model from IoT-based smart device's app source code and uses model checking to detect safety and security issues [56].

5.2.2 Dynamic smart home safety assurance frameworks

This type of safety assurance techniques is achieved to overcome the limitations of design phase SA. They usually extend and overlap with design-time approaches. They also called dynamic or system-control-based safety assurance frameworks because their effect continues through the whole system life cycle. These methods build upon the design-time approaches and use them to their benefit. In this sense, we emphasize they are not solely run-time paradigms, see **Figure 3**, so they (most of them) are a combination of static and dynamic approaches.

Run-time safety assurance approaches try to provide, according to Denny et al. [57] "safety cases for through-life safety assurance" via introducing a harmonic cooperation between design-time and run-time approaches. So, the system in the runtime phase has to pass through sense-detect-asses-mitigate loop, see **Figure 2** to overcome any safety-related design pitfalls.

The sense-detect-asses-mitigate loop is summarized as follows:

- *Sense:* is the first step to grasp operational data and passing them to next step,
- *Detect*: is the second step and it is responsible for detecting the suspicious system behavior during running,
- *Asses*: third step that receives its input from Detect step to evaluate and to decide the status the system entering (safe, vulnerable, hazardous, or consequent). Accordingly, the safety engine will decide whether to go to Mitigate step or to continue running the system,
- *Mitigate*: this step is the last choice the safety engine may call for.

Mitigation is not always available or possible but if it possible then can be implemented in two main ways:

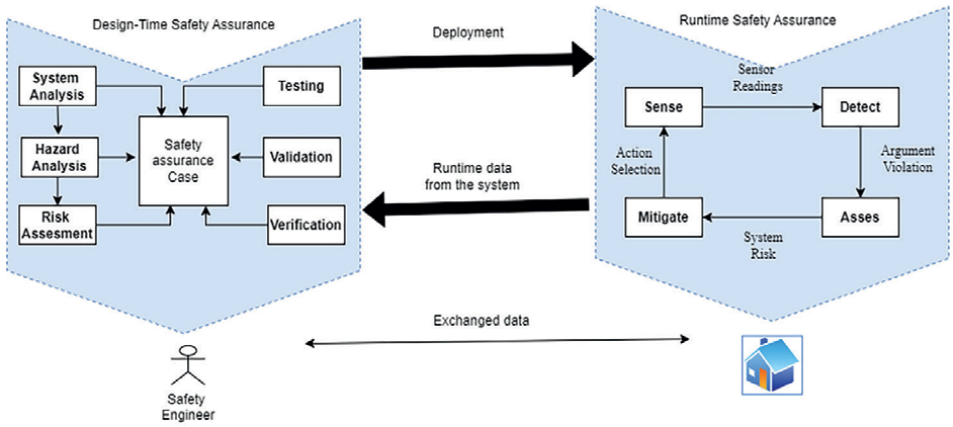

Figure 3.
Safety assurance in design time and runtime, adopted from [42], with some modifications.

1. The first is by blocking the system in its current status and trying to figure out how to return the system back to its previous status.

2. The second way is by augmenting the system with a verified safety wrapper (a backup version of the safety engine or safety controller) that can take control of the SH in order to avoid violations of formal safety properties.

Under unsafe operating conditions or system faults, the decision logic switches from the main safety engine to the safety controller to maintain safety [58]. Main drawbacks of this type of SA frameworks are:

1. They use quantitative data which put a processing burden on SH devices,

2. Hard to grasp, comprehend, and implement,

3. Have to depend on already-made design artifact.

Main advantages of this type of SA frameworks are:

1. Most of them provide risk mitigation strategies,

2. They capture the dynamic nature of the SH.

Most of the dynamic safety assurance methods make use of design-time safety risk management processes as a blueprint for the run-time safety assurance processes. Then, the remedies suggested during the run-time operation of the system are used as feedback information to enhance design-time safety assurance, see **Figure 4**.

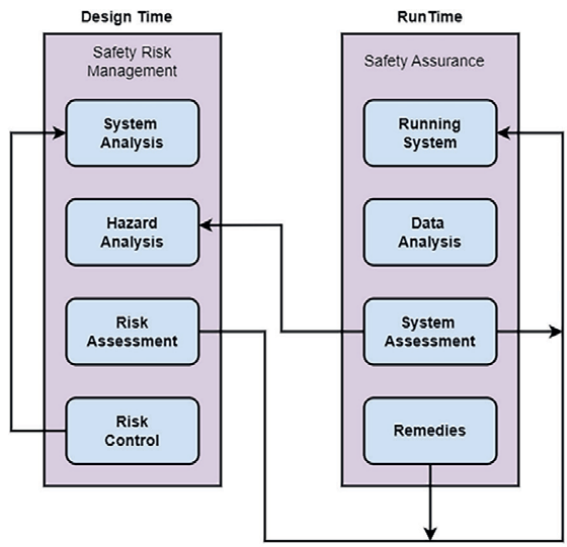

Figure 4.
The FAA framework for system safety management adopted from [59] with some modifications. It shows static and dynamic security assurance.

Several dynamic safety assurance methods have been proposed by smart home community's researchers previously. A prominent approach for dynamic safety assurance has been suggested: it involves automatically identifying trigger-action programming rule semantics from the source codes of IoT applications [60]. In this approach, the meaning of the source code rules is derived, and the TAPInspector builds these rules along with rule interactions as a finite state machine (FSM).

A new method for detecting dynamically, named PATRIoT, has been suggested [61]. While the IoT app is running, PATRIoT observes the entire system's interaction. If PATRIoT detects a breach of any defined security/safety policy, it will restrict the app's activity.

Celik et al. [62] designed a dynamic policy enforcement tool, IOTGUARD, which supports safety and security violations detection and enforcement in IoT apps. IOTGUARD directly blocks unsafe and undesired states in an individual app and multi-app environments.

A proposal for automatically transforming the IoT system into a Linear Hybrid Automata (LHA) model is introduced [63]. The inspection procedure aims to identify safety and security breaches and offer repair recommendations to its customers. The way they carry out the verification involves three main steps: Utilizing Linear Hybrid Automata Automatic Modeling to create LHA models, Examining the reachability of LHA models to assess the system against both positive and negative states through path-oriented checking, and finally, generating suggestions for fixing any verification issues.

A method called IoTBox, which relies on data mining techniques, is suggested for analyzing data produced by sandboxes in IoT systems [64]. Next, it is employed to identify alterations in behavior within IoT systems. A sandbox created based on rules from a smart home will yield minimal false positives and generates rules that can be checked by a human user. IoTBox detects the context in which actions are carried out and generates rules to identify actions that are either missing or in violation. Nevertheless, the requirement for human engagement renders this method somewhat infeasible and inappropriate for independent IoT SH systems.

The researchers in [65] use a combination of FTA and fuzzy neural networks in aquaculture IoT systems. They formulate an intelligent method for fault diagnosis. In their approach, the FTA is manually constructed for each component of the system, and later, they use a rule-based style on rules extracted from the FTA to be used as inputs for the fuzzy neural network so that to train the relationship model between fault safety violations and faults.

McCall et al. [66] present an approach called SAFETAP for safe trigger-action programming (TAP) paradigm. They use it to create automation rules that are triggered based on some condition and perform an action as a result. It is a dynamic safety assurance technique for SHs. SAFETAP is a rule-based analyzer that works on top of symbolic model checking (SMC) algorithms for checking TAP rules for any violations of the properties.

An approach called IotCom is introduced for analyzing hidden and unsafe interaction threats in IoT-based smart home through composition analysis [67]. IotCom utilizes path-sensitive static analysis to create an inter-procedural control flow graph (ICFG) for every application and then employs a graph abstraction method to represent the behavior pertaining to the connected devices in the app as a behavioral rule graph (BRG). BRG creates rules by connecting the triggers, actions, and logical conditions of each control flow in IoT applications.

A dynamic testing method for IoT physical interaction discovery called IOTSAFE is introduced [68]. IOTSAFE creates physical models of devices based on identified interactions, to forecast potential risky situations and prevent unsafe device states. A prototype of IOTSAFE was developed and integrated into the SmartThings platform.

Another dynamic approach is presented [69], named HOMEGUARD. HOMEGUARD is a system designed for IoT platforms in the form of apps to identify and address Cross-App Interference (CAI) risks. An automation semantics extracting module is created for IoT apps. The meanings of various IoT applications are examined together to assess how they interact and identify potential CAI risks.

6. Conclusion

The widespread adoption of IoT-based smart home systems in both private and public sectors necessitates that safety assurance must be given appropriate consideration to avoid the catastrophic consequences of underlying malfunctions. This study presents a different point of view by surveying major static and dynamic safety assurance approaches and highlighting their advantages and drawbacks. The study focuses on the independent relationship between cybersecurity and safety by explaining the major approaches used for safety assurance in isolation of security factor. This is achieved via splitting safety assurance in smart home from security issues and to draw a slim border between runtime and design-time safety assurance. The study's assessment method is based on the analysis of works published in previous authentic, peer-reviewed, and famous scientific conferences and journals indexed in relevant scientific databases, in addition to survey studies for distinguished related works. The study provides a larger scope in the field of IoT safety than previous studies; hence, it can e productively used by future researchers in the smart home safety assurance and give convenient and deeper comprehension and guidance for the IoT-based smart home topic's researches and professionals.

For future outlook, the study recommends that the design-time and runtime methods must be amalgamated into general methods to provide ongoing safety assurance guarantee. Moreover, safety can be affected by a combination of devices working together at the same time, so a holistic view of the system when designing safety guards must be taken.

Author details

Mouiad Al-Wahah* and Auhood Al-Hossenat
University of Thi-Qar, Thi-Qar, Iraq

*Address all correspondence to: mouiad-al@utq.edu.iq

References

[1] Iten R, Wagner J, Zeier RA. On the identification, evaluation and treatment of risks in smart homes: A systematic literature review. Risks. 2021;**9**(6):113

[2] Arcaini P, Bombarda A, Bonfanti S, Gargantini A, Riccobene E, Scandurra P. The ASMETA approach to safety assurance of software systems. In: Logic, Computation and Rigorous Methods: Essays Dedicated to Egon Börger on the Occasion of His 75th Birthday. Cham: Springer International Publishing; 2021. pp. 215-238

[3] European Commission. Commission implementing regulation (EU) No 1035/2011. Official Journal of European Union. 2011:19

[4] Dezfuli H, Allan B, Smith C, Stamatelatos M, Youngblood R. NASA System Safety Handbook. Volume 1, System Safety Framework and Concepts for Implementation. USA, Washington D.C: National Aeronautics and Space Administration; 2011

[5] Miller JE, Brown EK. The Cambridge Dictionary of Linguistics. 1st ed. Cambridge University Press; 2013. DOI: 10.1017/cbo9781139049412. Available from: https://www.cambridge.org/core/product/identifier/9781139049412/type/book

[6] Miorandi D, Sicari S, De Pellegrini F, Chlamtac I. Internet of things: Vision, applications and research challenges. Ad Hoc Networks. 2012;**10**(7):1497-1516

[7] Choudhary S, Mena G. Internet of things: Protocols, applications and security issues. Procedia Computer Science. 2022;**215**:274-288

[8] Hasan AK, Munam AS, Khan S, Ali I, Imran M. Perception layer security in internet of things. Future Generation Computer Systems. 2019;**100**:144-164

[9] Zou Z, Li K-J, Li R, Wu S. Smart home system based on ipv6 and zigbee technology. Procedia Engineering. 2011;**15**:1529-1533

[10] Wang P, Chaudhry S, Li L, Li S, Tryfonas T, Li H. The internet of things: A security point of view. Internet Research. 2016;**26**(2):337-359

[11] Tiwary A, Mahato M, Chidar A, Chandrol MK, Shrivastava M, Tripathi M. Internet of things (IoT): Research, architectures and applications. International Journal on Future Revolution in Computer Science & Communication Engineering. 2018;**4**:23-27

[12] Sethi P, Sarangi SR. Internet of things: Architectures, protocols, and applications. Journal of Electrical and Computer Engineering. 2017. pp. 1-25

[13] Kakkar L, Gupta D, Saxena S, Tanwar S. IoT architectures and its security: A review. In: Proceedings of the Second International Conference on Information Management and Machine Intelligence, Jaipur, India; 24-25 July 2020. pp. 87-94

[14] Hassija V, Chamola V, Saxena V, Jain D, Goyal P, Sikdar B. A survey on IoT security: Application areas, security threats, and solution architectures. IEEE Access. 2019;**7**:82721-82743

[15] Aswale P, Shukla A, Bharati P, Bharambe S, Palve S. An overview of internet of things: Architecture, protocols and challenges. Information

and Communication Technology for Intelligent Systems. 2019;**1**:299-308

[16] Ammar M, Russello G, Crispo B. Internet of things: A survey on the security of IoT frameworks. Journal of Information Security and Applications. 2018;**38**:8-27

[17] Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M. Internet of things: A survey on enabling technologies, protocols, and applications. IEEE Communications Surveys and Tutorials. 2015;**17**(4):2347-2376

[18] Tucic M, Pavlovic R, Papp I, Saric D. Networking layer for unifying distributed smart home entities. In: 2014 22nd Telecommunications Forum Telfor (TELFOR). IEEE; 2014. pp. 368-371

[19] SR e, S. System theoretic process analysis: A literature survey on the approaches used for improving the safety in complex systems. In: Information Systems for Industry 4.0: Proceedings of the 18th Conference of the Portuguese Association for Information Systems. Cham: Springer International Publishing; 4 May 2019. pp. 97-114

[20] Xing L. Reliability in internet of things: Current status and future perspectives. IEEE Internet of Things Journal. 2020;7(8):6704-6721

[21] Hammi B, Zeadally S, Khatoun R, Nebhen J. Survey on smart homes: Vulnerabilities, risks, and countermeasures. Computers & Security. 2022;**117**:102677

[22] Balakrishnan S, Vasudavan H, Murugesan RK. Smart home technologies: A preliminary review. In: Proceedings of the 6th International Conference on Information Technology: IoT and Smart City; 29 December 2018. pp. 120-127

[23] Abdulhamid A, Kabir S, Ghafir I, et al. An overview of safety and security analysis frameworks for the internet of things. Electronics. 2023;**12**(14):3086

[24] Qureshi KN, Abdullah AH. A survey on intelligent transportation systems. Middle-East Journal of Scientific Research. 2013;**15**(5):629-642

[25] Bakirtzis G, Carter BT, Elks CR, Fleming CH. A model-based approach to security analysis for cyber-physical systems. In: 2018 Annual IEEE International Systems Conference (SysCon). IEEE; 23 April 2018. pp. 1-8

[26] Abuserrieh L, Alalfi MH. Security and Safety Verification in IoT Apps. 2023 IEEE International Conference on Software Maintenance and Evolution (ICSME), Bogotá, Colombia. 2023. pp. 601-605. DOI: 10.1109/ICSME58846.2023.00080

[27] Zalewski J. IoT safety: State of the art. IT Professional. 2019;**21**(1):16-20

[28] Ismail Y. Introductory chapter: Internet of things (IoT) importance and its applications. In: Internet of Things (IoT) for Automated and Smart Applications. London, UK: IntechOpen; 27 November 2019

[29] Leveson NG. Engineering a safer world: Systems thinking applied to safety (engineering systems). Cambridge: MIT Press; 2011

[30] Haefner K, Ray I. ComplexIoT: Behavior-based trust for IoT networks. In: 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). IEEE; 12 December 2019. pp. 56-65

[31] Kriaa S, Pietre-Cambacedes L, Bouissou M, Halgand Y. A survey of

approaches combining safety and security for industrial control systems. Reliability Engineering & System Safety. 2015;**139**:156-178

[32] Wolf M, Serpanos D. Safety and security in cyber-physical systems and internet-of-things systems. Proceedings of the IEEE. 2018;**106**:9-20

[33] Zhou J, Cao Z, Dong X, Vasilakos AV. Security and privacy for cloud-based IoT: Challenges. IEEE Communications Magazine. 2017;**55**:26-33

[34] Matheu SN, Hernandez-Ramos JL, Skarmeta AF. Toward a cybersecurity certification framework for the internet of things. IEEE Security and Privacy. 2019;**17**:66-76

[35] Al-Swed WR, Al-Wahah MA. Trust as a pre-defense step for IoT authorization. Journal of Physics: Conference Series. 2021;**1963**(1):012172

[36] Boeckl KR, Fagan MJ, Fisher WJ, Lefkovitz NB, Megas KN, Nadeau EM, et al. Considerations for managing internet of things (IoT) cybersecurity and privacy risks. NISTIR. 2019;**8228**:1-34

[37] Li J, Zhao Z, Li R, Zhang H. AI-based two-stage intrusion detection for software defined IoT networks. IEEE Internet of Things Journal. 2019;**6**:2093-2102

[38] Sohal AS, Sandhu R, Sood SK, Chang V. A cybersecurity framework to identify malicious edge device in fog computing and cloud-of-things environments. Computers & Security. 2018;**74**:340-354

[39] Minoli D, Occhiogrosso B. Blockchain mechanisms for IoT security. Internet of Things. 2018;**1**:1-13

[40] Al-Wahah M, Farkas C. Context-aware IoT authorization: A dynamic and adaptive approach. In: 13th International Conference for Internet Technology and Secured Transactions (ICITST-2018). 2018. pp. 64-72

[41] Hartsell C, Mahadevan N, Ramakrishna S, Dubey A, Bapty T, Johnson T, et al. Model-based design for CPS with learning-enabled components. In: Proceedings of the Workshop on Design Automation for CPS and IoT; 15 April 2019. pp. 1-9

[42] Ramakrishna S. Dynamic safety assurance of autonomous cyber physical systems [PhD dissertation] Vanderbilt University. 2022

[43] Leveson NG. Software safety in embedded computer systems. Communications of the ACM. 1991;**34**(2):34-46

[44] Abdulkhaleq A, Wagner S, Leveson N. A comprehensive safety engineering approach for software-intensive systems based on STPA. Procedia Engineering. 2015;**128**:2-11

[45] Lemaire L, Lapon J, Decker BD, Naessens V. A SysML extension for security analysis of industrial control systems. In: 2nd International Symposium for ICS & SCADA Cyber Security Research 2014 (ICS-CSR 2014); 2 September 2014. pp. 1-9

[46] Nordmann A, Munk P. Lessons learned from model-based safety assessment with SysML and component fault trees. In: Proceedings of the 21th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems; 14 October 2018. pp. 134-143

[47] Sommerville I. Software. Engineering. 9th ed. Pearson Education,

Inc., publishing as Addison-Wesley. 2011. p. 18. ISBN-10.137035152

[48] Ericson CA. Hazard Analysis Techniques for System Safety. Hoboken, New Jersey: John Wiley & Sons; 2015

[49] Misra KB. Handbook of Performability Engineering. London: Springer Verlag; 2008. DOI: 10.1007/978-1-84800-131-2

[50] Wongvises C, Khurat A, Fall D, Kashihara S. Fault Tree Analysis-Based Risk Quantification of Smart Homes.

[51] Saeed F, Paul A, Rehman A, Hong WH, Seo H. IoT-based intelligent modeling of smart home environment for fire prevention and safety. Journal of Sensor and Actuator Networks. 2018;**7**(1):11

[52] Onofrio R, Piccagli F, Segato F. Failure mode, effects and criticality analysis (FMECA) for medical devices: Does standardization foster improvements in the practice? Procedia Manufacturing. 2015;**3**:43-50

[53] Lee J, Wu F, Zhao W, Ghaffari M, Liao L, Siegel D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. Mechanical Systems and Signal Processing. 2014;**42**(1-2):314-334

[54] Nguyen DT, Song C, Qian Z, Krishnamurthy SV, Colbert EJ, McDaniel P. IotSan: Fortifying the safety of IoT systems. In: Proceedings of the 14th International Conference on Emerging Networking EXperiments and Technologies; 4 December 2018. pp. 191-203

[55] Liang CJ, Karlsson BF, Lane ND, Zhao F, Zhang J, Pan Z, et al. SIFT: building an internet of safe things. In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks. 2015. pp. 298-309

[56] Celik ZB, McDaniel P, Tan G. Soteria: Automated {IoT} safety and security analysis. In: 2018 USENIX Annual Technical Conference (USENIX ATC 18). 2018. pp. 147-158

[57] Denney E, Pai G, Habli I. Dynamic Safety Cases for through-Life Safety Assurance. In: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering; 16 May 2015. Vol. 2. IEEE; pp. 587-590

[58] Bak S, Manamcheri K, Mitra S, Caccamo M. Sandboxing controllers for cyber-physical systems. In: 2011 IEEE/ACM Second International Conference on Cyber-Physical Systems. 12 April IEEE; 2011. pp. 3-12

[59] Federal Aviation Administration [Online]. Advisory Circular (AC120-92A), 2021. Available from: https://www.faa.gov/documentLibrary/media/AdvisoryCircular/AC%20120-92A.pdf

[60] Yu Y, Liu J. TAPInspector: Safety and liveness verification of concurrent trigger-action IoT systems. IEEE Transactions on Information Forensics and Security. 2022;**17**:3773-3788

[61] Yahyazadeh M, Hussain SR, Hoque E, Chowdhury O. Patriot: Policy assisted resilient programmable iot system. In: Runtime Verification: 20th International Conference, RV 2020, Los Angeles, CA, USA, 6-9 October, 2020, Proceedings. Springer International Publishing; 2020. pp. 151-171

[62] Celik ZB, Tan G, PD MD. IoTGuard: Dynamic enforcement of security and safety policy in commodity IoT. In: NDSS Symposium. San Diego, CA, USA. 24-27 February 2019. ISBN 1-891562-55-X. 2019

[63] Bu L, Xiong W, Liang CJ, Han S, Zhang D, Lin S, et al. Systematically ensuring the confidence of real-time home automation IoT systems. ACM Transactions on Cyber-Physical Systems. 2018;**2**(3):1-23

[64] Kang HJ, Sim SQ, Lo D. Iotbox: Sandbox mining to prevent interaction threats in IoT systems. In: 2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST). IEEE; 12 April 2021. pp. 182-193

[65] Chen Y, Zhen Z, Yu H, Xu J. Application of fault tree analysis and fuzzy neural networks to fault diagnosis in the internet of things (IoT) for aquaculture. Sensors. 2017;**17**(1):153

[66] McCall M, Shezan FH, Bichhawat A, Cobb C, Jia L, Tian Y, et al. SAFETAP: An Efficient Incremental Analyzer for Trigger-Action Programs. Pittsburgh, PA, USA: Carnegie Mellon University; Rep. 14792271, 2021

[67] Alhanahnah M, Stevens C, Bagheri H. Scalable analysis of interaction threats in iot systems. In: Proceedings of the 29th ACM SIGSOFT International Symposium on Software Testing and Analysis; 18 July 2020. pp. 272-285

[68] Ding W, Hu H, Cheng L. IOTSAFE: Enforcing safety and security policy with real IoT physical interaction discovery. In: Network and Distributed System Security Symposium. 2021

[69] Chi H, Zeng Q, Du X, Yu J. Cross-app interference threats in smart homes: Categorization, detection and handling. In: 2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). IEEE; 29 June 2020. pp. 411-423

Chapter 5

The Impact of 5G-Enabled Edge-Cloud Services on Energy Facilities in Industry 4.0

Nikolaos Tzanis, Eleftherios Mylonas, Panagiotis Papaioannou, Christina (Tanya) Politi, Alexios Birbas, Christos Tranoris, Spyros Denazis, Ioannis Moraitis, Alex Papalexopoulos, Anna Tzanakaki and Jesús Gutiérrez Terán

Abstract

As the energy sector undergoes a profound digital transformation, demanding a fusion of resilience, efficiency, and cutting-edge technology, 5G technology emerges as a beacon, promising not just enhanced connectivity but a holistic transformation of how we conceive and manage energy infrastructure. This work aims to provide an in-depth exploration for experts in the energy domain, unraveling the innovative aspects of 5G through the demonstration of important achievements and results of the Horizon 2020 5G Infrastructure Public Private Partnership (5G-PPP) Phase-3 5G-VICTORI's Project and its trial results on the impact of 5G technology in an energy facility located in the city of Patras, Greece.

Keywords: edge-cloud continuum, industrial IoT, 5G, private networks, virtualization, geofencing, artificial intelligence

1. Introduction

The digitization of the industrial sector has led to the arrival of the first Industry 4.0 services, which resulted in enabling new markets within the sector and creating new opportunities based on modern industrial procedure paradigms. These paradigms are the basis for the so-called Smart Factory vertical industry, which integrates and benefits from emerging technologies like Internet of things (IoT), artificial intelligence (AI), edge/cloud computing, etc., and aims to bring a new era in the way the industrial sector collects, processes and utilizes data.

From a network requirement point of view, the Smart Factory vertical can be split into different services or applications, presenting different requirements. Maintenance activities require the support of low-cost, energy-efficient sensors planted in a distributed and heterogeneous infrastructure. Security and operation services ask for low latency trip signals and high bandwidth for closed circuit

television (CCTV). 5G technology is expected to integrate and deliver services able to support these diverse applications simultaneously.

5G technology offers a suite of features tailored to enhance operational efficiency and data security in industrial environments [1]: *(1) Enhanced Network Performance*, offering superior network performance for different types of applications through massive machine type communication (mMTC), ultra-reliable low latency communication (URLLC) and enhanced mobile broadband (eMBB). This performance boost enables seamless connectivity crucial for industrial processes; *(2) Slicing Support for Uninterrupted Performance*, 5G's support of slicing ensures a dedicated lane for critical applications, guaranteeing they meet their key performance indicators (KPIs) consistently. This capability enables the concurrent support of diverse applications with varying requirements, ensuring operation even in the face of fluctuating background traffic. The concept of network slicing was first introduced in Release 15 of the Third Generation Partnership Project (3GPP) 5G specification [2], while Release 17 defined a mechanism in order to support multiple service level agreement (SLA) requirements and introduced energy efficiency KPIs [3]; *(3) Co-location of 5G Core (5GC) Functions and Application Functions (AFs)*, which significantly reduces latency in communication, fostering real-time responsiveness crucial for industrial processes and concurrently enhances data privacy by keeping critical functions close to the source; *(4) Secure Isolation with 5G Non-Public Network (NPN)*, 5G's support of NPNs ensures that sensitive industrial data remains safeguarded, addressing the need for secure communication within industrial ecosystems. Moreover, it enables the monitoring and configuration of the communication network internally, allowing for quick response in case of incident or demand for re-configuration due to the inability to meet specific KPIs. The concept of NPN was first introduced in Release 17 of the 3GPP 5G specification [3]; *(5) Flexible Deployment of Edge Processing Services*, 5G technology facilitates the seamless deployment of new services demanding edge processing. This agility allows industrial setups to swiftly integrate cutting-edge solutions, empowering them to adapt to evolving technological demands with ease; and, finally, *(6) Cost Reduction*, from 5G's ability to use commercial-of-the-shelf (COTS) servers rather than dedicated network equipment to execute network functions, the support of multiple services on a single infrastructure and the ability to push intelligence at the edge, contributes to significant cost savings for industrial entities. The use of wireless 5G sensors in the "last mile" of deployment also contributes to cost reduction (e.g., at substation level).

This work aims to present an in-depth exploration designed for experts in the energy field. It delves into the innovative dimensions of 5G, illustrating notable accomplishments and findings within the context of the Factories of the Future use case in the 5G-VICTORI project [4]. 5G-VICTORI is a Horizon 2020 5G-PPP Phase-3 project whose main goal is to showcase large-scale field trials for advanced use case verification in commercial environments deploying 5G infrastructures in support of a number of vertical industries, specifically the Factories of the Future, Transportation, Energy, Media verticals.

The objective of the 5G-VICTORI Factories of the Future use case is to demonstrate that, by leveraging the unique features of 5G technology and 5G-VICTORI architecture [5–7], novel Industry 4.0 applications with different requirements can be sufficiently supported in a private 5G network deployment. This use case intends to demonstrate two different scenarios:

- The application of mMTC-banded IoT architectures for preventive maintenance and monitoring of the factory assets.

DOI: http://dx.doi.org/10.5772/intechopen.1005514

- The support of uRLLC and eMBB applications for real-time monitoring, security, and automation in an industrial environment.

This chapter is structured as follows: Section 2 introduces the Factories of the Future use case and the three different services it comprises namely the Operation, Maintenance, and Facilities Security service. Section 3 describes the Operation service's details and results. In Section 4, the Maintenance service's details are analyzed, whereas Section 5 deals with the Facilities Security service. In Section 6, we present the challenges we faced during the 5G-VICTORI project. Lastly, we conclude our work in Section 7, where the outcomes of the Factories of the Future use case are summarized.

2. Use case details and architecture

The trial takes place at the Independent Power Transmission Operator (IPTO or ADMIE in greek) facilities located in Patras, Greece, which consist of two sites, separated by 4 km of sea, and are electrically interconnected via a high-voltage (HV) 150 kV submarine power cable. Each site serves as termination point for the submarine HV power cable and comprises one control room and several sensors for the monitoring of the termination equipment status at this site. The facilities are interconnected with the Patras5G testbed [8], located at the University of Patras (UoP), via millimeter-wave (mmWave) links provisioned by Intracom Telecom.

Patras5G offers a cloud platform that can host core network components and mobile edge computing (MEC) deployments. By leveraging the 5G-VICTORI architecture, 5GC functions can be co-located with AFs at an edge datacenter (e.g., Autonomous Edge—located at the ADMIE site) or the cloud according to the requirements of the service. **Figure 1** depicts the architecture of this use case

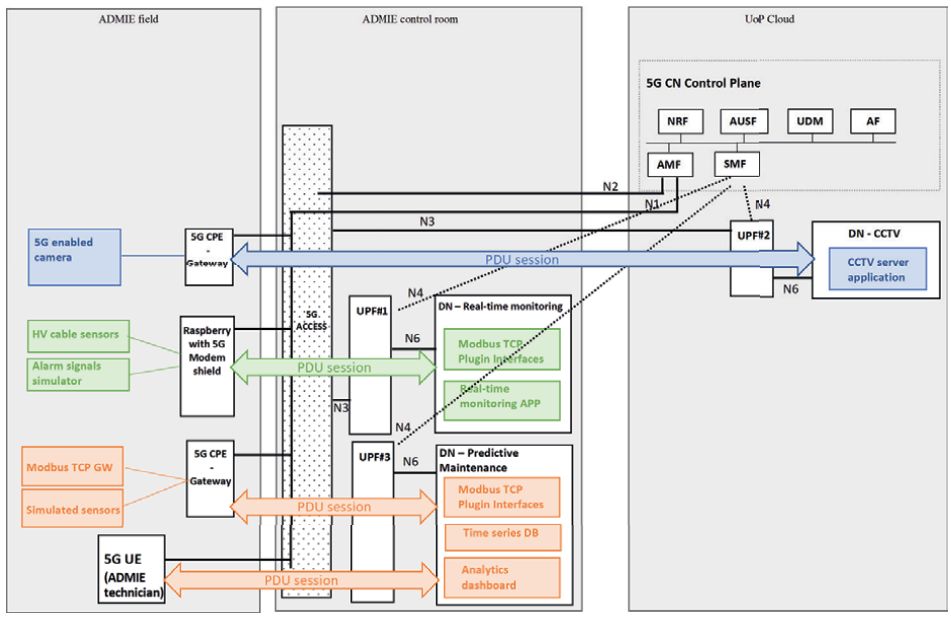

Figure 1.
High-level architecture of the factories of the future use case.

and highlights the placement of application functions and 5GC functions to support its applications. More details of the architecture description can be found in 5G-VICTORI's deliverable D3.6 [9].

The applications that are demonstrated refer to operation, maintenance, and facilities security domains [9, 10]. The following sections describe the objective of each application and the high-level deployment methodology, focusing on 5G innovative aspects, and present results and valuable findings for the energy operator.

3. Operation: real-time monitoring of HV power cable

Operation-related applications refer to applications involved in the real-time monitoring of the system and are characterized by guaranteed low latency and high-reliability requirements. In the Real-Time Monitoring of HV Power Cable, measurements of active power (MW), reactive power (MVar), and current (A) from the primary and secondary windings of the transformer at each site are collected and compared in real time. Frequency (Hz) measurements are also collected on one side of the cable via a high-frequency measurement device.

3.1 Methodology

Measurements collection is performed through Modbus transmission control protocol (TCP) clients developed for the project. Legacy sensors are connected to the private 5G network via 5G customer premises equipment (CPE) gateways. By combining voltage, current and frequency measurements, we can classify power events in the area to automate control system events and faults. This is feasible by utilizing wavelet signal processing to capture abrupt changes in the grid's voltage, current and frequency signals. Since only frequency measurements are time-stamped, to provide valid results, the latency difference of the measurements must be maintained below a specific threshold. Traditionally, this is accomplished by providing dedicated fiber connections, leading to expensive and inflexible legacy solutions. For this trial, legacy sensors are enhanced with 5G capabilities and are connected to a wireless NPN, which can meet the specified KPIs imposed by the real-time monitoring application, while providing a more economical and easily expandable/maintainable solution. To maintain measurements of latency difference below a specific value, network latency and network jitter (variation in network latency) should meet specific upper bound values.

Collected measurements during the trials are visualized on an online dashboard, which is depicted in **Figure 2**.

The future objective for this application is to be able to feed an automated controller with timely results. Toward this objective, processing must be performed as close to the sensors as possible.

3.2 Innovative aspects/results

3.2.1 User plane at premises (low latency, increased privacy)

To support the above requirements, the real-time monitoring application—and supporting 5GC functions are hosted at the ADMIE site. This choice leads to a significant reduction in end-to-end (E2E) latency and increases privacy as measurements never leave ADMIE premises.

Figure 2.
Operation: online dashboard for visualization of critical information.

E2E latency	13.7 ms (min)	38.5 ms (average)	86.8 ms (max)
Network latency difference	0.0002 ms (min)	0.7516 ms (average)	2.085 ms (max)
Jitter	< 0.0001 ms (min)	0.272 ms (average)	0.3 ms (max)
Sensor datarate	54 Kbps		
Measurements requests (transactions)	1000 requests		

Table 1.
Operation: field results.

HV monitoring application presents 38.5 ms (mean) end-to-end latency. The latency difference between the two sites has a mean value of 0.75 ms and a max of 2.08 ms. This means that the service can estimate adequately the steady state and dynamic state of the network (dynamic phenomena of 10^{-2} s).

3.2.2 Specific slice (guaranteed low latency, increased reliability)

As it is shown in **Figure 1**, the application uses a dedicated slice for its data network, meaning that the performance of this critical application is ensured and independent of background traffic produced by other services. This increases application reliability. Network delay and delay variation (jitter) are minimized, leading to better synchronization of packets between the sites and a lower rate of obsolete packets.

The results of the operation service are summarized in **Table 1**.

4. Maintenance: sensor data collection for preventive maintenance

This application uses a large number of measurements originating from different types of sensors to monitor the health of the actual HV power cable. The submarine cable is impregnated with a low-viscosity insulation fluid (oil) to increase its operating dielectric stress, working temperature and current carrying capacity. Oil pressure,

oil temperature and other relevant measurements are continuously collected and processed to estimate the health of the cable.

4.1 Methodology

The application uses at its core the Unified IoT Orchestration Platform (UiTOP) (see **Figure 3**), which is an industrial IoT platform offered by Intracom Telecom [11]. This application is compatible with many industrial protocols, provides rich dashboards and customizable alarms, and follows a microservices architectural approach. The application does not have strict latency requirements, as it is used for predictive maintenance analytics. On the contrary, it must be able to handle, store and analyze vast amounts of data originating from different types of sensors. Moreover, it should be accessed only by the maintenance crew inside the facilities. This is accomplished through the geofencing (-or location-based service provisioning) feature.

4.2 Innovative aspects/results

4.2.1 Virtualization, standard APIs support

Both network function and IoT platform follow a containerized approach, thus increasing scalability—it is easy to go from 10 sensors to 1000 sensors—and portability. Moreover, the adoption of microservices architectural approach and the use of REST APIs

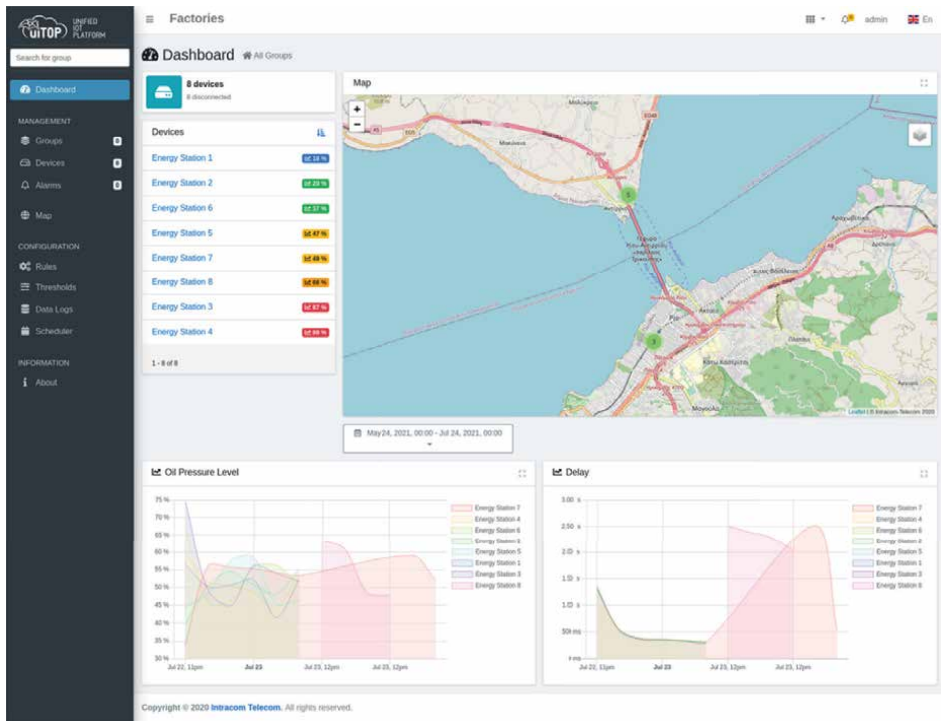

Figure 3.
Maintenance: UiTOP dashboard also shows live measurements collection delay.

DOI: http://dx.doi.org/10.5772/intechopen.1005514

for information exchange facilitates the integration of third-party tools for advanced analytics not supported by the platform (e.g., machine learning to predict equipment aging).

4.2.2 Geofencing

Geofencing is a virtual boundary that, in our case, is accomplished through the 5G radio domain. More precisely, the network can be configured to provide different access levels to the same application, according to the base station that the user is registered. This means that a user with 5G user equipment (UE) can access the application (or specific features of an application) only when he is inside the facilities served by the specified base station.

Geofencing offers two valuable features to the preventive maintenance solution:

1. *Enhanced privacy*: Even if an intruder finds credentials for the application, he will not be able to access specific information if he is not physically inside the facilities.

2. *Information presented on the application's dashboard is facilities-specific*: This facilitates the inspection process. Inspector crews, who often travel to remote locations, do not have to search for equipment codes in long lists, but measurements of interest are automatically shown according to their location.

To showcase the advantages of the proposed 5G-VICTORI architecture, two identical gNodeBs (gNBs) are deployed at the Patras5G cloud facility (emulating the Central Offices of ADMIE) and at the ADMIE Rion facility (facility environment), which are featured in **Figure 1**.

Location-based service provisioning is realized via the selection of different data networks according to the gNB in which the UE is registered. **Figure 4** depicts the UiTOP dashboard accessed from a UE connected to gNB2 at ADMIE premises. Since UE is connected to gNB2 it uses Data Network #2 (*DN* #2) that has access to UiTOP service running on Autonomous Edge. On the contrary, when the UE is registered to the NPN through gNB1 at the University campus, it uses *DN* #1 , and cannot access the specific service.

Figure 4.
UiTOP dashboard accessed by UE only at ADMIE premises (right), custom dashboard created for a specific type of measurements (left).

Sensor datarate	~0 Kbps (min)	11 Kbps (average)	54 Kbps (max)
Total datarate	~0 Mbps (min)	1.07 Mbps (average)	5.40 Mbps (max)
Packet loss ratio	~0		
# of sensors	100 sensors		
Measurements requests (transactions)	6000 requests		

Table 2.
Maintenance: field results.

As shown in **Figure 4**, inspection workers have access to information regarding the health of the HV power cable, but they also have access to information regarding the communication network status (the dashboard also presents network latency and bandwidth). NPN deployments offer complete control to the vertical industry (in this case, the power transmission operator) to monitor and (re)configure/expand the network according to their needs without relying on external network providers.

4.2.3 NPN deployment (increased monitoring, increased flexibility)

Continuous monitoring of telecommunication networks is crucial for critical infrastructures. NPN deployments also offer complete control to the vertical industry. The ability to reconfigure/expand/maintain the telecommunication network without relying on third-party network operators increases the reliability, security and privacy of the critical infrastructure, as it is easier to identify bottlenecks, prioritize critical applications and steer traffic when application needs change.

The results of the maintenance service are summarized in **Table 2**.

5. Facilities security: smart CCTV surveillance service for industrial environments over 5G

Industrial infrastructures must be monitored, not only for the security of the facilities themselves but also for the technical personnel to ensure their physical well-being. To this end, CCTV monitoring of facilities over 5G will provide a solution to this problem by collecting and processing live video feeds when technical personnel is present or an event occurs, while not compromising other Industry 4.0 applications running in the background.

The application does not have strict latency requirements since it focuses on throughput and sending multiple, stable, high-quality video streams from the edge (ADMIE facilities at Rion) to the cloud (UoP facilities) without interruption. In order to demonstrate this and test the application in real working conditions of industrial environments over 5G and 5G NPNs, the application is deployed in parallel with the two aforementioned Factories of the Future services. Application isolation and guaranteed quality-of-service (QoS) is achieved via the network slicing feature of 5G.

5.1 Methodology

A static ultrahigh definition (UHD) CCTV camera and a 5G-enabled mobile surveillance robot (**Figure 5**) are used for collecting live video feeds from the ADMIE

Figure 5.
Facilities security: 5G-enabled mobile surveillance robot in action.

Figure 6.
Facilities security: capture from the CCTV camera and analyzed by the intruder detection service, accessed by UoP.

facilities. The CCTV camera inside the HV cable control room monitors the industrial equipment installed there while the mobile surveillance robot patrols outside and near dangerous high-voltage areas. The video streams from the CCTV camera feed an AI-empowered intruder detection algorithm running at the UoP cloud (**Figure 6**). The video collected from the robot is streamed to a web server hosted at UoP, which can be accessed as-is by the facilities' supervisor. The service runs isolated from the others while no QoS degradation is detected.

5.2 Innovative aspects

5.2.1 Network slicing

Provision of dedicated network slice for guaranteed video streaming, regardless of background traffic, for both static camera and mobile surveillance robots.

5.2.2 Support of AI

AI has become a standard tool in developing new industrial services since it offers numerous benefits. In the Factories of the Future use case, this is demonstrated via the use of an AI-powered facilities monitoring and intruder detection algorithm running in the cloud, which minimizes human intervention with regard to facility security. The inclusion of AI is crucial not only for the mitigation of work accidents in such hazardous environments but also offers an effective solution for facilities located at remote sites, since there is no need for physical presence. The AI solution presented in this scenario is based on a pre-trained region-based convolutional neural network (RCNN) in TensorFlow for human detection, which runs in the UoP cloud. The application receives live video from the interior of the HV cable control room, analyzes it, and notifies the user (facilities' supervisor) in case someone enters the facilities. The user is then able to identify if this is a scheduled visit or an intrusion incident and act accordingly. More advanced intrusion detection techniques for critical facilities monitoring—without humans in the loop—could include biometrical identification algorithms [12], which were not investigated during the 5G-VICTORI project.

5.2.3 Use of 5G-enabled surveillance robot

The use of a 5G-enabled surveillance robot gives great opportunities. 5G-enabled drones, often used in power grids, are suitable for overhead power line inspection spread at remote locations, but they suffer at identifying faults near the ground (e.g., the base of power towers) or hard-to-reach spots inside the facilities. In case of an incident, a surveillance robot can be used to identify if it is safe for the personnel to reach the area and provide a useful insight into the situation. In collaboration with swarms of drones, surveillance robots can provide a unified surveillance solution for power utilities.

The results of the Facilities Security service (as summarized in **Table 3**) prove the seamless CCTV service provisioning over 5G with full isolation from other running

Video streaming latency (plus application processing overhead)	0.9 s (min)	2.32 s (average)	5.01 s (max)
CCTV camera datarate	6.5 Mbps (min)	9.33 Mbps (average)	12.78 Mbps (max)
5G-enabled mobile surveillance robot camera datarate	0.4 Mbps (min)	1.5 Mbps (average)	2.5 Mbps (max)
Aggregated datarate	7.3 Mbps (min)	11.03 Mbps (average)	15 Mbps (max)

Table 3.
Facilities security: field results.

services and without introducing performance errors. In addition, they show smart facility monitoring services, that is, intruder detection, inside power utility facilities, which are a special case of the industrial environment due to their harsh conditions, for example, high-voltage and electromagnetic interference. Finally, the ability to stream high-quality video stream from a moving surveillance robot to a remote location improves the safety of first responders in the case of event, and the security of the overall facility.

6. Challenges raised during the project

The process of designing, developing, deploying, and validating the aforementioned services and the necessary setups for both lab and field trials during the 5G-VICTORI project was faced with a lot of hardships. One of the most critical issues

Figure 7.
Installation of mmWave antenna equipment linking ADMIE and UoP sites.

Figure 8.
Field trials for factories of the future.

during the project was the interconnection of the ADMIE facilities with the Patras5G testbed via a mmWave link (**Figure 7**). This task was essential for the development of the services as well as the preparation of the final demonstration actions, and it proved extremely difficult due to the management and orchestration issues it raised, since different teams with different expertise backgrounds had to work together. In addition to this, the hostile environment of power utilities resulted in permanent equipment damage after a thunderstorm in the vicinity of the facilities, which again demanded new installation works at the site. Furthermore, the project's great ambition to provide real network slicing for applications of diverse requirements and deployment flexibility resulted in much development work, eventually leading to a fully configurable 5G network deployment flow. Through this flow, features like geofencing were possible. Last but not least, adapting industrial protocols into a 5G environment and enabling legacy sensors to be integrated into a 5G network required the development of appropriate protocol adapters, which came to be crucial for the successful demonstration of the services (**Figure 8**).

7. Conclusions

The inclusion of 5G technology in industrial environments and more specifically critical infrastructures is a crucial step toward the realization of the Factories of the Future concept. Through 5G, flexible network architectures that satisfy both low latency and high-bandwidth application requirements are possible. In addition, thanks to the network slicing feature of 5G, complete isolation between different applications of diverse required QoS is achieved without performance penalties and regardless of background noise. These features, along with the reduction of installation costs due to the use of standard COTS equipment for private 5G network deployments, plus the wireless nature of the sensors, are especially critical from

the operators' point of view since they make 5G a viable and powerful solution. Furthermore, advanced data privacy is possible through the geofencing feature, which enables location-based service provisioning and can further safeguard industries from cyber threats. Lastly, 5G networks are able to support the requirements of emerging AI technologies that will further contribute to the development of future smart services for the industry sector.

Acknowledgements

This work was supported by the H2020 European Project 5G-VICTORI under Grant 857201.

Abbreviations

5G-PPP	5G Infrastructure Public Private Partnership
IoT	Internet of things
AI	artificial intelligence
CCTV	closed circuit television
mMTC	massive machine type communication
uRLLC	ultra-reliable low latency communication
eMBB	enhanced mobile broadband
KPI	key performance indicator
3GPP	Third Generation Partnership Project
SLA	service level agreement
5GC	5G Core
AF	application function
NPN	non-public network
COTS	commercial-of-the-shelf
HV	high-voltage
UoP	University of Patras
mmWave	millimeter-wave
MEC	mobile edge computing
CPE	customer premises equipment
E2E	end-to-end
UE	user equipment
gNB	gNodeB
DN	data network
QoS	quality-of-service
UHD	ultrahigh definition
RCNN	region-based convolutional neural network

Author details

Nikolaos Tzanis[1,2†], Eleftherios Mylonas[1,2*†], Panagiotis Papaioannou[2†], Christina (Tanya) Politi[2†], Alexios Birbas[2†], Christos Tranoris[2†], Spyros Denazis[2†], Ioannis Moraitis[1†], Alex Papalexopoulos[3†], Anna Tzanakaki[4†] and Jesús Gutiérrez Terán[5†]

1 Department of Research, Technology and Development, Independent Power Transmission Operator (IPTO) S.A., Athens, Greece

2 Department of Electrical and Computer Engineering, University of Patras, Patras, Greece

3 Ecco International Inc., San Francisco, CA, USA

4 National and Kapodistrian University of Athens, Athens, Greece

5 IHP Leibniz-Institut Für Innovative Mikroelektronik, Frankfurt (Oder), Germany

*Address all correspondence to: e.mylonas@admie.gr

† These authors contributed equally.

IntechOpen

References

[1] Empowering Vertical Industries through 5G Networks—Current Status and Future Trends. Available from: https://5g-ppp.eu/wp-content/uploads/2020/09/5GPPP-VerticalsWhitePaper-2020-Final.pdf [Accessed: January 23, 2024]

[2] 3GPP TS 23.501 Version 15.9.0 Release 15. Available from: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.09.00_60/ts_123501v150900p.pdf

[3] 3GPP TS 22.261 Version 17.11.0 Release 17. Available from: https://www.etsi.org/deliver/etsi_ts/122200_122299/122261/17.11.00_60/ts_122261v171100p.pdf

[4] 5G-VICTORI Website. Available from: https://www.5g-victori-project.eu/ [Accessed: January 23, 2024]

[5] 5G-VICTORI Deliverable D2.2. Available from: https://www.5g-victori-project.eu/wp-content/uploads/2020/05/2020-05-21-5G-VICTORI_D2.2_v1.0.pdf

[6] 5G-VICTORI Deliverable D2.3. Available from: https://www.5g-victori-project.eu/wp-content/uploads/2021/07/2021-06-11-5G-VICTORI_Deliverable_D23_v1.0_submitted_EC.pdf

[7] 5G-VICTORI Deliverable D2.4. Available from: https://www.5g-victori-project.eu/wp-content/uploads/2022/05/2022-04-11-D2.4-5G-VICTORI-end-to-end-reference-architecture.pdf

[8] Patras5G. Available from: https://wiki.patras5g.eu/ [Accessed: January 23, 2024]

[9] 5G-VICTORI Deliverable D3.6. Available from: https://www.5g-victori-project.eu/wp-content/uploads/2023/02/2022-06-17-5G-VICTORI_D3.6-Final-Test-Cases-for-Energy-and-Factories-of-the-Future_v1.0.pdf

[10] 5G-VICTORI Deliverable D3.5. Available from: https://www.5g-victori-project.eu/wp-content/uploads/2022/05/2021-07-31-5G-VICTORI-D3.5-Prel-Test-Cases-for-Energy-and-Factories.pdf

[11] Unified IoT Orchestration Platform. Available from: https://www.intracom-telecom.com/en/products/telco_software/IoT/IoT_orchestration_Platform.htm [Accessed: January 23, 2024]

[12] Minaee S, Abdolrashidi A, Su H, et al. Biometrics recognition using deep learning: A survey. Artificial Intelligence Review. 2023;**56**:8647-8695. DOI: 10.1007/s10462-022-10237-x

Chapter 6

Service Provision and Price Strategies in Edge Computing

Xiulong Liu, Xiaoyi Tao, Sheng Chen and Xin Xie

Abstract

This chapter of research explores the strategic deployment of computing services at the edge of the network to optimize social welfare and system performance. It involves building models for where and how to place services within edge computing systems, considering factors such as demand, cost, and usability. Additionally, this chapter investigates dynamic pricing frameworks that allow for real-time pricing adjustments to operate resource allocation efficiently. These strategies aim to balance the economic objectives of service providers, such as profit maximization and cost minimization, with the quality of service delivered to users. The goal is to develop pricing and service placement mechanisms that are both economically beneficial and capable of meeting the latency, computational, and energy consumption demands of edge applications.

Keywords: service provision, request allocation, economic model, social welfare, price strategies

1. Introduction

Edge computing is a technology that shifts data processing from centralized data centers to locations closer to the data source, such as smartphones, factory sensors, or vehicles. This approach offers several benefits: it reduces the time and distance for data transmission, enabling faster data processing results; alleviates the load on central servers; and enhances privacy and security since data do not have to be transmitted over long distances.

Now, I will explain the connection between edge computing and several application areas shown in **Figure 1**:

- Industrial Internet of Things (IIoT): In industrial settings, numerous sensors and devices collect data (such as temperature, pressure, and speed). Edge computing enables rapid processing of these data near these devices, allowing real-time monitoring and optimization of factory operations without the need to send large amounts of data to remote servers.

- Wireless Sensing: In wireless sensing technology, devices detect and understand their surroundings through wireless signals. Edge computing can process this

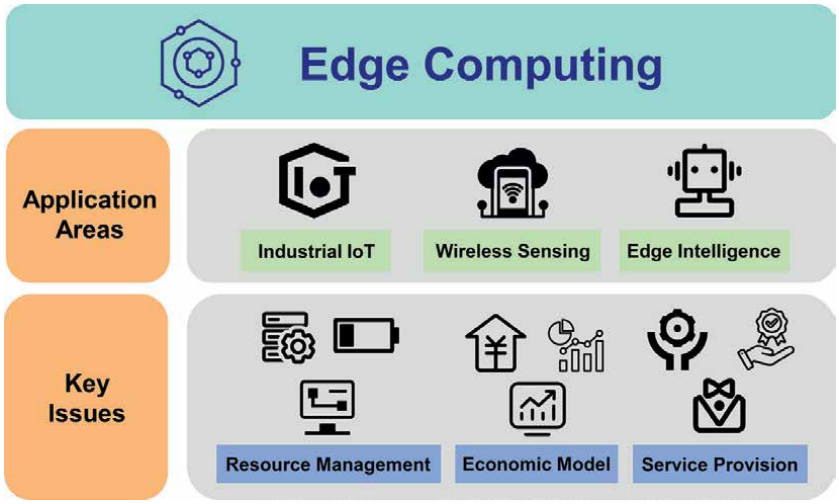

Figure 1.
A description of application and issue in edge computing.

information immediately at the edge, supporting rapid responses in applications such as smart homes and security monitoring.

- Edge Intelligence: This refers to the application of Artificial Intelligence (AI) within an edge computing framework. For example, in a smart factory, AI can be used for immediate analysis of data collected from machine sensors, enabling quick decision-making, such as predicting equipment failures in advance, thereby reducing downtime.

Within these applications, three key issues of edge computing are highlighted: resource management, economic modeling, and service provision. Here, we will use straightforward language to delve into these seemingly complex concepts.

- *Resource management: the magic kitchen of edge computing*

  Imagine edge computing as a magical kitchen. In this kitchen, our primary task is to ensure that each chef (representing data processing devices) has ample ingredients (data) and tools (computing resources) for efficient operation. Proper allocation of ingredients leads to faster and better-quality dish preparation (processing results). This underscores the importance of resource management in edge computing, ensuring the optimal utilization of each resource.

- *Economic model: the profit rule of edge computing*

  Let us discuss the economic model. In our magical world, this is likely to establish a set of rules that ensure the profitability and operational sustainability of our magical kitchen. It is like finding the perfect recipe that not only attracts customers (users) to savor our culinary creations but also ensures profitability for our kitchen. This "recipe" must balance the costs (the expenses of running edge computing) and the revenues (fees paid by users for services), ensuring fairness to users and profitability for operators.

- *Service provision: the secret to customer satisfaction in edge computing*

  Here, we arrive at service provision. The main goal is to ensure every customer receives the service they expect, much like ensuring every dish in the magical kitchen satisfies the customer's taste. In edge computing, this means users can swiftly and reliably receive the data processing services they need. Additionally, this service must be secure, just as restaurants ensure the safety and hygiene of the food they serve.

In general, the importance of resource management, economic modeling, and service provision is to ensure the successful operation of a magical kitchen. Resource management ensures an efficient cooking process, the economic model secures the kitchen's profitability and sustainable growth, while high-quality service provision guarantees customer satisfaction. When these three elements coexist harmoniously, the magic of edge computing is maximized, offering users a swift, safe, and efficient data processing experience. This is similar to a successful restaurant, where delicious dishes, reasonable pricing, and excellent service are key to retaining customers.

Therefore, when discussing edge computing, this section is not only talking about technology but also a strategy encompassing efficient resource utilization, the formulation of sensible business models, and the provision of optimal user experiences. This is the essence of the magic of edge computing and a significant reason for its emergence as a future technological trend.

The remainder of this chapter is organized as follows: it begins with an introduction to applications at the edge, followed by a detailed analysis of service provision and pricing strategies from the perspectives of resource management, economic modeling, and service delivery. In detail, Section 2 introduces applications of fog computing and human sensing. Section 3 details methods for resource management and optimization. Section 4 discusses the economic model and pricing strategies. Section 5 provides an analysis of service provision and networking. Section 6 provides a conclusion for the chapter.

2. Technological innovations and applications

Recently, researchers have explored developments in Mobile Edge Computing (MEC) and Industrial IIoT. It highlights new methods and systems that improve internet transit for MEC providers. Today, MEC is vital in telecommunications because it places computing power near the data source, reducing delay, and saving bandwidth. Section 2 also discusses fog computing – an extension of cloud computing that further reduces data processing time. Additionally, Section 2 covers wireless human sensing technologies, which allow remote monitoring of human activities and conditions. These technologies largely explain the importance of edge computing applications.

In the era of big data, cloud computing has significantly enhanced daily life and technology development. However, cloud computing's centralized nature and potential for latency issues in time-sensitive services, along with security risks like privacy breaches during data center failures or network disconnections, are notable shortcomings. Fog computing emerges as a complementary solution, extending computation to the cloud's edge. It primarily employs edge devices, like routers and local servers, offering enhanced security and sustainability. Unlike cloud computing's

large centralized resources, fog computing utilizes numerous distributed, renewable fog nodes, each with modest computational power. This distributed approach reduces data center workload, improves efficiency, and reduces costs. Despite its advantages, fog computing also brings new security challenges, especially with the interconnectedness of IoT devices, raising concerns like personal data theft and exposure. Du details a novel differential privacy-based query model for fog data centers, focusing on enhancing data privacy in heterogeneous fog computing environments. The model integrates Laplacian noise to safeguard data, balancing privacy protection, and data utility. This approach effectively resists privacy attacks, crucial for large-scale datasets in fog computing. Both fog and MEC deal with data processing closer to the source, where privacy and security are significant. Applying this model could significantly improve data privacy and security in MEC mirroring the benefits observed in fog computing [1].

Wireless human sensing is a crucial aspect of human-computer interaction, as it allows computers to recognize and understand human activities and even emotions. Numerous endeavors have been undertaken to tackle this issue through various wireless technologies, including WiFi, RFID, Bluetooth, Radar, and Zigbee. Each of these sensing technologies possesses distinct characteristics and advantages, rendering them appropriate for particular application contexts. For instance, WiFi-based systems are adept at non-intrusive human sensing, whereas RFID-based systems excel in identifying individuals in environments with multiple persons. A comprehensive review is provided to assist users in comprehending the range of available wireless human sensing technologies, thereby facilitating an informed decision regarding the most fitting solution for their requirements. This review specifically delves into promising human sensing applications and classifies them into three categories: vital sign monitoring, gesture recognition, and activity recognition [2]. The advent of wireless human sensing applications has yielded a multitude of profound benefits for society, particularly in realms such as healthcare, gesture recognition, and the recognition of human activities. Within a Smart Wireless Human Sensing (SWHS) system, the dynamics of human movement apply a significant influence on the propagation of wireless signals. Using the resulting changes in signal characteristics, human activity can be identified by non-invasive methods.

The importance of healthcare has become increasingly apparent in recent years, especially in the context of an aging global population, which poses a major challenge to existing healthcare infrastructure. Increased attention has been given to routine healthcare as it is directly linked to the serious health problems prevalent among older persons. One such health issue is sleep apnea, a serious medical condition characterized by interrupted breathing during sleep. More than half of the world's population aged 65 and above suffer from a sleep disorder. The prevalence of sleep apnea syndrome is estimated to be between 20 and 40%, a proportion that increases with age. The manifestations of sleep apnea can be observed through a variety of physical symptoms, including episodes of shortness of breath or apnea, increased blood pressure, potential heart complications, skin discoloration, and even loss of consciousness.

Traditional diagnostic methods, such as polysomnography (PSG), while widely used in clinical settings, come with significant inconvenience and financial burden, especially for older patients who often require care and supervision at home. To address these challenges, developing and implementing respiratory monitoring systems using ubiquitous wireless signals has emerged as a promising alternative. These systems are non-invasive and provide a more convenient and accessible way to monitor respiratory health.

3. Resource management and optimization

Resource management and optimization in edge computing involve efficiently allocating and utilizing computational resources (such as processing power, storage, and bandwidth) at the edge of the network, close to the data sources. This process aims to enhance application performance, reduce latency, and manage the constraints of edge devices, which often have limited resources compared to centralized cloud computing resources. The objective is to optimize the use of limited and distributed resources in edge computing environments to meet the demands of applications and services, ensuring they run effectively and efficiently. Section 3 will discuss the feature extraction, performance, and cost in heterogeneous edge systems. The advantages of the following methods can be concluded. The feature extraction problem in edge systems benefits for network limitations [3, 4], the performance problem helps the tradeoff between energy consumption and latency [5], and the load balance problem is conducive to latency-sensitive applications [6]. However, these methods only focus on partial performance elements, network power consumption, latency, etc. In most cases, resource allocation and optimization problems need to consider multiple factors.

In the rapid expansion of the Internet of Everything, network bandwidth constraints are a major challenge in transferring large amounts of data to cloud or edge servers. To alleviate this problem, a common strategy is to pre-process image data on devices before uploading it to edge servers, thereby reducing network traffic and bandwidth pressure. A key aspect of the process is the extraction of discriminant features from images, which is crucial for recognition tasks. However, due to the lack of image label information necessary for feature extraction and the fluctuating availability of resources on these devices, this task is challenging on mobile devices.

Mobile devices often run multiple applications at the same time, and frequent startup and shutdown of applications can affect resource availability. Therefore, it is important to develop discriminative image feature extraction methods, based on current mobile device resources, to reduce network traffic while maintaining high recognition accuracy. Ding proposes a method to solve this problem [3]. The detailed steps are as follows: after uploading the pre-processed image data to the edge server, the server processes the data and returns the label information of the most similar image to the mobile user. The main challenge is to create an effective feature extractor, called Extractor E, which can identify key discriminant features. These features are critical because they significantly affect the quality of service (QoS) of moving image recognition, including recognition accuracy and response time. To solve this problem, the authors introduce the discriminant feature extraction (DFE) algorithm. The DFE algorithm is designed to generate an extractor E that can extract a minimal but efficient discriminant feature set from image data, improving recognition accuracy and shortening response time by reducing network traffic and the number of matched features. This is achieved by building a new similarity function that preserves the intra and inter class structure of the image dataset and by introducing a tradeoff parameter that balances these structures [3].

The second major challenge is dynamically selecting extractors based on a mobile device's fluctuating resources. To address this issue, Ding also introduces the Nested Discriminative Feature Extraction (NestDFE) algorithm. This algorithm segments the extractor E into several sub-extractors, collectively constituting a multi-capacity extractor. Notably, this multi-capacity extractor occupies the identical memory space as the original extractor E. This efficiency is achieved as each sub-extractor of lesser capacity shares parameters with and is embedded within, a sub-extractor of greater

capacity, thereby obviating the need for additional memory space. Consequently, the NestDFE algorithm permits mobile devices to dynamically choose the most suitable sub-extractor without necessitating substantial memory space consumption [3].

Traditionally, these processes have relied on mobile cloud computing, where users upload images to cloud servers for retrieval results. However, these methods often suffer from long network transmission delays. To solve this pr oblem, Wang proposes a method leveraging MEC, which allows mobile users to interact with edge servers that are closer to them than cloud servers, thereby reducing transmission latency [4]. By facilitating faster data processing and reduced response times, MEC plays a key role in various fields such as the Internet of Things, e-healthcare, and autonomous vehicles. One challenge with existing MEC-based image retrieval methods is the need to extract a large number of features from the image, which must then be uploaded to a cloud server. Since extracting features from image datasets stored in the cloud is isolated, the process can be inefficient and result in poor retrieval accuracy.

Wang introduces a new cloud-guided feature extraction method for mobile image retrieval. In this approach, a cloud server learns the projection matrix P from its dataset of labeled images. Subsequently, this matrix P is utilized to extract discriminant features from images, thereby creating a low-dimensional feature dataset. The edge server then applies matrix P to the image, and the resultant features are uploaded to the cloud server for label recognition. This technique notably diminishes network traffic and simultaneously enhances retrieval accuracy. A prototype system was implemented to validate this approach, and extensive experiments in a real MEC environment are conducted. The results show a remarkable reduction in network traffic by 93% and an improvement in retrieval accuracy by 6.9% compared to existing state-of-the-art image retrieval methods in MEC [4].

Another work focuses on optimizing energy efficiency in MEC while ensuring performance [5]. This work addresses the challenge of balancing energy consumption and performance in computation offloading tasks in MEC environments. The authors propose an energy-minimizing optimization problem and solve it using the Karush-Kuhn-Tucker (KKT) conditions. The solution involves a request offloading scheme based on energy consumption and bandwidth capacity. Numerical results demonstrate that the proposed offloading scheme offers better energy consumption and delay performance compared to local computing and complete offloading methods. The research contributes to the field by providing an effective method for computation offloading in MEC, which is crucial for mobile users who require lower energy consumption and better performance for their tasks. This research is particularly relevant for mobile users who seek to balance energy consumption with effective performance in their computational tasks within MEC environments [5].

Edge computing has emerged as the foremost method for delay-sensitive applications, strategically placing compute and storage resources at the network's edge. Its fundamental function is to efficiently manage data transmission between the cloud downlink and the terminal uplink and to organize these data effectively at the edge, laying the groundwork for subsequent data analysis and processing. This raises a critical question: How can data be efficiently organized, stored, and retrieved at the edge? To address this, a variety of methods have been devised for establishing data storage and retrieval services at the edge, encompassing structured, unstructured, and hybrid approaches. However, research in heterogeneous edge environments regarding data storage and retrieval services is still lacking, particularly in load balancing of edge data stores. The principal aim of edge computing is to optimally utilize edge node resources, including storage, bandwidth, and CPU, to satisfy user requirements.

Given the limited resources of individual edge nodes, cooperation among nodes becomes essential. A common strategy is to distribute the workload evenly across edge nodes, thereby optimizing resource utilization. Chen introduces the w-strategy, a novel load-balancing technique that uses weighted Voronoi graphs to achieve optimal distribution across heterogeneous edge nodes [6]. Using the software-defined networking (SDN) paradigm, the proposed approach supports data distribution based on virtual space-distributed hash tables (DHTS). Such heterogeneity presents additional complexities in load distribution. The w-strategy comprises two key components. Firstly, utilizing the SDN paradigm, this work maps nodes and data within the edge network to virtual-space coordinates in the SDN control plane. The w-strategy then divides the load area of each node to achieve balanced distribution, based on the virtual-space location of nodes and data. Secondly, w-strategy accounts for edge node heterogeneity by assigning weights to nodes according to their storage and computational capabilities and subsequently distributing loads based on these weights. The main challenge of the w-strategy lies in integrating these two aspects: achieving even load distribution while considering node capabilities. Evaluation results demonstrate that the w-strategy outperforms existing methods, such as greedy routing for edge data (GRED) and Chord, by enhancing resource utilization efficiency by an average of 20% [6].

4. Economic models and pricing strategies

Compared with cloud computing data center systems, the resources in edge computing systems are very limited, which leads to increased competition for resources among users who expect high-quality services. Inefficient resource allocation decisions may lead to low service quality, high energy consumption, and high operating costs. At present, most of the work is focused on the design of resource allocation algorithms in edge computing systems. The pricing mechanism has a significant effect in the competitive use of resources. The decentralized distribution of edge nodes, the heterogeneity of resource requirements, and the competition among users for high-quality services make resource allocation and pricing in edge computing systems a challenging problem.

4.1 Pricing mechanisms in edge computing

The distributed nature of edge nodes, the heterogeneity of resource demands, and the competition among users for high-quality services render the allocation and pricing of resources in edge computing systems a challenging issue. Price mechanisms have some advantages with this challenge including:

1. In the edge environment, providers often aim to maximize profits while satisfying the needs of edge users, making the protection of providers' interests the primary goal of edge resource management. Pricing models, such as profit maximization or cost minimization, can effectively achieve this goal.

2. The edge environment encompasses a variety of roles belonging to different groups, such as end-users, edge service providers, network service providers, and edge brokers, each with their distinct objectives and demands, such as cost, profit, and utility, and facing various constraints like budget limits and resource capacities. Pricing mechanisms can quantify these different goals and attempt to solve them.

3. More and more new edge applications and computing modes, such as big data analysis and distributed machine learning frameworks, require a significant amount of computing and bandwidth resources. Pricing mechanisms, such as bandwidth congestion perception pricing, can regulate users' demand for resources and maximize resource utilization.

Often, these roles in the edge systems have conflicting goals, making economic models and pricing strategies highly effective for the allocation and management of edge resources. More specifically, through negotiation and game mechanisms among these roles, under given constraints, incentive and pricing mechanisms can provide the best solutions for self-interested parties. These works mainly consider the edge and node pricing model and seldom take the chain of cloud-edge-node situation.

4.2 Dynamic pricing strategies for edge computing systems

For an individual Edge Service Provider (ESP), its primary objective is to obtain revenue through the provision of services to users. A prerequisite for its participation in an edge federation is that it can reduce the costs associated with constructing and maintaining edge nodes. In the motivation shown in **Figure 2**, each edge node is associated with a base station, enabling users to access the network via these base stations. It is assumed that a user has agreed with ESP 1 in edge 1, meaning the user can only access the network through ESP 1's base station. Initially, the user's service is located on edge 1. As a consequence, the user is required to pay a fee to ESP 1. Upon

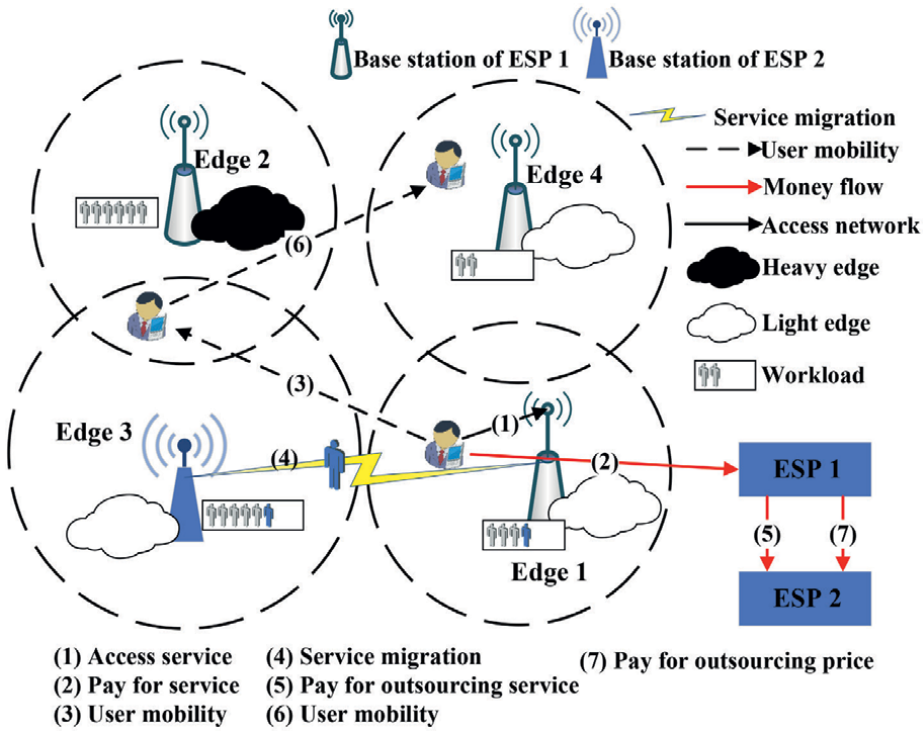

Figure 2.
Motivation for dynamics pricing for edge systems.

leaving the service area of edge 1 and entering the combined service areas of edge 2 and edge 3, the system must decide whether to migrate the service in response to the user's movement. Due to edge 2 being at full capacity, it cannot provide efficient service, necessitating that the user's service remain on edge 1. However, the increasing distance from the user to edge 1, where their service is hosted, would inevitably lead to higher communication delays, significantly impacting the user's Quality of Service (QoS). Thus, the optimal solution is for ESP 1 to form a federation with ESP 2, granting ESP 1 the right to utilize a portion of ESP 2's resources. Consequently, the system can migrate the user's service to edge 3, which is closer to the user with some additional migration delay. To maintain the stability of the federation, ESP 1 compensates ESP 2 for servicing the customer's request. As the user moves beyond the service area of edge 3 to that of edge 4, the system weighs the migration delay against the communication delay to decide whether to migrate the service. If the service is not migrated, ESP 1 continues to compensate ESP 2. This analysis demonstrates that the edge federation expands the service range and capabilities of an ESP, enabling it to leverage idle resources for greater profit. Essentially, the process involves redirecting user service requests through the federation to the most cost-effective edge node for processing. The dynamic cooperative optimization of service placement and pricing by network edge nodes can enhance both the user experience and ESP revenue, achieving a win-win situation.

Dynamic pricing usually sets the price of resources within a small time frame and can quickly respond to changes in resource demand. Especially in situations where user information is unknown, the system can adjust prices based on the real-time scarcity of resources to pursue high returns from resource providers and efficient allocation of resources. Han [7] studied the dynamic pricing scheme of idle vehicle resources under the condition that the arrival of vehicles is random and unknown and minimized the cost of the entire edge computing system under the premise of guaranteeing the quality of service. Xue [8] utilized D2D technology to fully utilize the computing resources of idle edge users to cope with network congestion caused by some hot spots. The author defines the ratio of computing resource prices to the service capabilities of candidate auxiliary users as the dynamic transaction price. The decision-making problems of resource allocation, task offloading, and candidate auxiliary user selection are defined as minimizing total energy consumption and total working user expenses. The original problem is decomposed using the alternating direction multiplier method to obtain the resource allocation and dynamic pricing scheme. Siew [9] also studied the problem of uneven resource allocation in multi access wireless networks based on this idea, and proposed two dynamic pricing mechanisms to achieve the supply of computing resources, which, respectively, improved social welfare and platform benefits. Wang [10] studied the dynamic pricing problem of drone resource supply in hotspot areas. Considering the differences in drone hover time and service capacity, the authors designed multiple drone deployment schemes for heterogeneous hotspot areas. By balancing the occurrence rate of users in hotspot areas and the flight distance of drones, the overall revenue of drones was improved. Although dynamic pricing can also address unknown user requests, it almost always considers short-term optimization goals and cannot provide overall service assurance. Therefore, Chen et al. proposed a real-time dynamic pricing strategy to optimize ESP's long-term goals [11, 12].

The primary goal in multiple-edge systems is to minimize Internet data transfer costs for small ESPs in the IIoT. Chen's research introduces the Sublessor framework [12], which reduces WAN data transmission costs by using nonprofit ESPs as neutral

brokers who charge only for technology and administration. This framework leverages the ability of MEC providers to share infrastructure and exchange network data locally, significantly lowering costs within the same region [12]. Chen's another work highlights challenges in edge computing, emphasizing the need for effective load evaluation on edge nodes due to the unpredictability of requests and timing [11]. This necessitates real-time, adaptive pricing decisions. Additionally, aligning pricing with resource allocation is complex, given user sensitivity to price fluctuations, which can impact customer satisfaction. Therefore, Chen proposes separating these processes to operate on different timescales. Furthermore, current dynamic pricing strategies often overlook long-term profit maximization, underscoring the importance of strategic, long-term planning despite the randomness of network requests.

4.3 Auction-based pricing strategies for edge computing systems

Auction is a popular form of transaction that efficiently allocates a seller's resources to a buyer at a competitive price in the marketplace. Many works study resource auctions in edge scenarios [13]. Hung [14] investigated live video streaming services bidding on resources of edge servers to improve the smoothness of live video streaming. The edge system allocates cache resources utilizing combinatorial clock auctions and uses dynamic planning to finalize the price and quota of cache resources, which improves the utility of the entire edge system. Jiao [15] considered a mobile blockchain scenario in which miners competed for edge resources to implement blockchain services and proposed a resource allocation method based on combinatorial auctions to improve social welfare. There is another work [16] aiming at the heterogeneous nature of edge users' demands and the competition for high service quality among multiple users. Two different resource allocation mechanisms are proposed, namely, an auction-based individually rational and envy-free allocation mechanism and a linear programming-based approximation algorithm. The algorithm does not guarantee the absence of jealousy, but the approximate solution can ensure that the difference with the optimal solution is controlled within a definite range.

Double Auction is designed for multi-edge and multi-user auction scenarios. Sun [17] proposes a credible and efficient algorithm rooted in a two-way auction model. It guarantees no-loss and employs dynamic bidding to match "mobile device-edge server" pairs while meeting local constraints. Multi-Round Auction is a flexible and efficient auction method, that facilitates online buyers to adopt different bidding strategies. Wang [18] firstly designs a set of non-competitive incentive mechanisms to encourage edge cloud to sell resources, while ensuring the QoS of users and the interests of edge cloud; then designed a multi-round auction mechanism for competitive environments to achieve matching of edge terminals and resources and maximize the profit of the edge cloud. Zhang [19] first proposed to fully exploit the vehicle computing resources in the parking lot as a supplement to the edge computing environment to provide low-latency services to the mobile users, and the ESP can shut down the idle edge servers to save energy and operation cost. The authors propose a multi-round auction to maximize the utility of resources, such as vehicle computing, which optimizes the service provisioning structure at the edge and achieves a win-win situation for both vehicles and ESPs. There is another work [20] recruiting edge server owners to carry the computing tasks of edge users for the cloud. Zhou's method uses the reverse auction method to select appropriate edge nodes to provide computing resources for edge users and minimize the cost of cloud service centers. Xu [21] studies the scenario of mobile users choosing trusted edge nodes to cache resources in a

campus environment and proposes a reverse auction-based method to select the most suitable edge node for users among the nodes that satisfy user requirements.

5. Service provisioning and networking

Addressing service provisioning and networking challenges by focusing on service quality, social welfare, and availability involves a subtle and integrated approach. Edge computing improves service quality by reducing latency and optimizing bandwidth, making it ideal for time-sensitive applications and efficient data transmission. However, maintaining consistent service quality across diverse edge environments poses challenges, including ensuring low energy consumption and managing the variability in computational resources [22, 23]. Edge systems can extend high-quality digital services to remote and underserved areas, contributing to greater accessibility and inclusivity. The challenge lies in ensuring that these benefits are equitably distributed and that edge systems are designed with social welfare in mind [24]. Edge computing also enhances system resilience and fault tolerance, ensuring service availability of virtualized network functions (VNF) even in the face of disruptions. The scalability of the service chain is another advantage, allowing the network to accommodate growing demands. However, ensuring high availability in diverse and potentially harsh operational environments requires sophisticated resource management and infrastructure resilience strategies [25, 26].

5.1 Service quality

Service quality in MEC is critical as it directly affects user satisfaction and the overall effectiveness of edge computing services. Ensuring low energy consumption and high reliability of data processing and transmission is the key to providing high-quality services. This is achieved by strategically placing services on edge nodes closest to users and optimizing network routing to reduce latency. In addition, consistent performance is critical even when demand or network conditions fluctuate. QoS metrics such as throughput, latency, and energy consumption are continuously monitored and managed to ensure that services meet expected standards.

Tao's work involves optimizing the placement of virtualized network functions (VNF) in multi-region mobile edge systems [22]. This work emphasizes balancing performance metrics such as latency and energy consumption. The authors propose a cost-driven VNF placement Edge (VPE) strategy, which effectively solves the VNF placement problem using a graph-cutting method. The VPE approach considers three cost types: communication, location, and energy consumption. This approach is superior to other approaches in terms of latency and energy consumption, providing a balanced solution for VNF placement that takes into account both user and system perspectives. This research contributes to improving edge system performance and cost efficiency, which is critical to supporting high-quality mobile user experiences in multi-regional edge environments. The proposed VPE solution uses the graph-cutting method to efficiently determine the location of low-cost and high-performance VNF. The study demonstrated the effectiveness of VPE in improving latency and energy consumption, with adjustable cost weights to meet the needs of different users [22].

For network issues, Deng addresses the critical need for efficient resource management, specifically in the context of sampling node selection and sampling duration allocation [23]. This work underscores the necessity of managing the constrained

resources in MEC systems with precision. The importance of meticulously determining the number of sampling nodes and judiciously allocating the duration of sampling is considered, especially considering the bandwidth constraints of packet sampling collectors. Subsequently, this paper delves into the application of Deep Reinforcement Learning (DRL) in resource allocation within wireless environments, referencing relevant studies. DRL, an amalgamation of Reinforcement Learning (RL) and deep learning, has demonstrated its efficacy in autonomously managing resource allocation and adapting to evolving time-based demands. For instance, a particular study developed a resource allocation strategy that accounts for the computational prowess of MEC servers and the cache capacities of base stations (BSs). Aiming to enhance network-wide flow-level sampling precision, the authors advocate for a strategy that involves adaptively choosing sampling nodes and dynamically formulating sampling techniques for each interval. This is achieved using the Simple Probabilistic Sampling (SPS) method. This approach aims to optimize network monitoring within the constraints of MEC systems' resources, addressing the challenges of achieving high accuracy in flow-level sampling [23].

5.2 Social welfare

Social welfare in the context of MEC involves making computing resources more accessible and benefiting a wider segment of society. This includes deploying edge computing solutions in a way that addresses the digital divide, ensuring that underserved communities also benefit from technological advances. It extends to creating fair resource allocation algorithms that distribute computing power and storage fairly between users and applications. By reducing the centralization of computing resources, MEC can contribute to a more balanced and inclusive digital ecosystem.

The existing research assumes that all edge nodes in a network are owned by a single Edge Internet Provider (EIP). However, the reality of the MEC environment is often characterized by the presence of multiple EIPs, such as AT&T in the United States, Mobile in China, and Bells in Canada. In such a diverse MEC landscape, two primary challenges arise, prompting our investigation into edge federation networks with multiple EIPs to realize the concept of edge federation. First, each EIP tends to build its own private edge computing environment, serving only the customers it contracts with its federation, while this approach limits each EIP to managing its edge nodes and serving only its customers. For EIP participating in edge federation, it is important to ensure that participants reduce costs when building and maintaining edge nodes while enhancing customer service capabilities. Determining how EIP in edge federated networks can gain greater social welfare is another major challenge. Chen solves the problem of service placement in a MEC network consisting of multiple EIP [24]. These EIPs are coordinated in a joint manner to maximize social welfare, a concept known as a federation. While seemingly simple, the task is complex due to several factors: the EIP must decide whether to service the request internally or outsource the request to another EIP, and determine the appropriate edge node to handle these tasks [24].

In addition, Edge Infrastructure Providers (EIPs) must decide on the number of requests to accept from fellow EIPs and allocate appropriate edge nodes for these requests. A critical question emerges: how can EIPs effectively schedule various user service requests (internal, incoming, and outsourced) to improve service performance? Furthermore, to augment the overall profit, each EIP should adopt strategic pricing for services offered to both customers and those from other EIPs within the federation. This involves adjusting prices in response to user QoS and request

capacity, aiming to optimize social welfare. Addressing these challenges, this paper introduces a model for horizontal collaboration among edge federations, encompassing edges from all EIPs. Initially, the model treats the service placement dilemma as a programming issue, to maximize social welfare. It then presents two dynamic pricing strategies for EIPs: one for setting standard prices for customers and another for determining insourcing prices for services from other EIPs [24].

5.3 Availability

Availability is another key factor in MEC service delivery. It refers to consistent and reliable access to computing resources and services. In an edge computing environment, where resources are distributed and can be subject to a variety of local conditions, maintaining high availability is challenging but essential. This includes designing robust systems that can withstand hardware failures, network outages, and fluctuations in user demand. Redundant strategies, such as replicating critical services across multiple edge nodes, and dynamic resource allocation, which can quickly respond to changes in demand or network conditions, are key to ensuring high availability.

With the rise of network function virtualization (NFV) and MEC, network service providers (NSPS) are increasingly outsourcing network functions (NFs) to MEC, as it provides greater scalability and flexibility for NFV deployment and maintenance. In this process, each user request is processed through a service function chain (SFC) consisting of multiple VNFs. These VNFs are software alternatives to traditional hardware-based middleware that are arranged in a specific order to respond to requests. However, unlike hardware, VNFs are less reliable due to potential software errors and host failures. To improve reliability, it is wise to incorporate redundancy in the primary VNF within the SFC, where the key issue is to determine the optimal MEC node to place each VNF and the number of backup instances required to meet the availability requirements of each SFC.

Consequently, Li proposes an availability-aware provisioning strategy for SFCs in the MEC environment, designated as APoS [25]. APoS main aim is to maximize the number of served requests while complying with the requirements and reliability standards of SFCs. In addition, APoS tackles two principal challenges: one, the effective mapping of both primary and backup Virtual Network Functions (VNFs) to fulfill the availability needs of SFCs, and two, the diminution of latency for user access to SFCs. The first challenge is conceptualized as an integer nonlinear programming (INLP) problem, considering the resource constraints of each MEC node. Addressing this NP-hard problem, Li employs a novel binary N-back search method to optimally position the primary and backup VNFs. Regarding the second challenge, the objective is to minimize the average delay for all requests within each time slot. To this end, this work introduces an online service switching (OSS) method. This method accounts for queuing, communication, and switching delays, providing an optimal solution underpinned by theoretical evidence. The evaluation of these methods using real-world datasets shows significant improvements over benchmarks. Specifically, the proposed method achieves an approximate 20% increase in request acceptance and up to a 30% reduction in delay on average [25].

Within NFV, SFC stands as a pivotal concept, denoting the sequence of VNFs that network traffic traverses from its source to its destination. The core challenge of NFV-enabled networks is SFC embedding, which requires optimally assigning VNFs to nodes and routing each stream so that it traverses the required SFC. Recent research has focused on various goals such as minimizing costs, ensuring availability,

and reducing latency. This approach can significantly reduce total latency, which is an inherent limitation of VNFs sequential processing in traditional SFC. Another key issue with NFV is that when separated from dedicated hardware, the vulnerability of network functions increases due to potential software failures or physical node failures. To address this issue, the researchers advocate adding backup VNF instances to enhance availability. The authors discuss a mixed-deployment SFC using sequential and parallel VNFs.

However, the dynamic, virtualized, and multiplexing nature of NFV also increases the risk of security vulnerabilities, presenting a larger attack surface. Current validation tools fall short in addressing the complexities of this environment, particularly given that existing guidelines predominantly cater to traditional, sequential Service Function Chains (SFCs), which are merely a subset of hybrid SFCs. As a result, agile and comprehensive validation techniques for hybrid SFCs are still uncharted territory. To address this deficiency, the authors have developed verification scheme for hybrid service function chain (vHSFC), a novel validation framework specifically designed for hybrid SFCs [26]. This framework enables enterprises to authenticate SFC execution in real time accurately. The vHSFC framework employs a streamlined, verified routing protocol to identify various SFC breaches and threats, including packet alterations and deviations from compliant SFC forwarding paths. This work details the creation and deployment of vHSFC prototypes using multiple containers, alongside rigorous testing with actual network traffic. The findings demonstrate that vHSFC is adept at ensuring consistent enforcement and detecting unforeseen breaches, all while sustaining a manageable level of system overhead [26].

6. Conclusion

In conclusion, this chapter has provided a comprehensive exploration of the strategic placement and pricing of computing services at the network's edge. The detailed analysis has demonstrated how optimal service placement in edge computing systems can significantly enhance social welfare and system performance. This involves careful consideration of demand, cost, and usability factors. Moreover, dynamic pricing frameworks have been investigated that support real-time pricing adjustments, ensuring efficient resource allocation and balancing economic objectives with quality of service. The findings suggest that these strategies not only are economically advantageous for service providers but also meet the stringent requirements of edge applications in terms of latency, computational power, and energy consumption. This chapter underlines the potential of edge computing to revolutionize service delivery in our increasingly connected world, offering a roadmap for service providers to optimize both their operational efficiency and user satisfaction.

Nomenclature

AI	artificial intelligence
MEC	mobile edge computing
IIOT	industrial Internet of Things
SWHS	smart wireless human sensing
PSG	polysomnography
QoS	quality of service

DFE	discriminative feature extraction
NestDFE	nested discriminative feature extraction
KKT	Karush-Kuhn-Tucker
VPF	virtualized network function
VPE	VNF placement edge
SDN	software defined networking
DHTs	distributed hash tables
MIP	mixed-integer programming
DRL	deep reinforcement learning
DREAM	dynamic pricing-based online control algorithm
RL	reinforcement learning
SPS	simple probabilistic sampling
BS	base station
EIP	edge internet provider
NSP	network service provider
NFV	network function virtualization
SFC	service function chain
INLP	integer nonlinear programming
OSS	online service switching

Author details

Xiulong Liu[1*†], Xiaoyi Tao[2†], Sheng Chen[1] and Xin Xie[1]

1 Tianjin University, Tianjin, China

2 Dalian Maritime University, Dalian, China

*Address all correspondence to: xiulongliu@tju.edu.cn

†These authors contributed equally.

References

[1] Du M, Wang K, Liu X, Guo S, Zhang Y. A differential privacy-based query model for sustainable fog data centers. IEEE Transactions on Sustainable Computing. 2017;**4**(2):145-155

[2] Liu Z, Liu X, Zhang J, Li K. Opportunities and challenges of wireless human sensing for the smart IoT world: A survey. IEEE Network. 2019;**33**(5):104-110

[3] Ding C, Zhou A, Liu X, Ma X, Wang S. Resource-aware feature extraction in mobile edge computing. IEEE Transactions on Mobile Computing. 2020;**21**(1):321-331

[4] Wang S, Ding C, Zhang N, Liu X, Zhou A, Cao J, et al. A cloud-guided feature extraction approach for image retrieval in mobile edge computing. IEEE Transactions on Mobile Computing. 2019;**20**(2):292-305

[5] Tao X, Ota K, Dong M, Qi H, Li K. Performance guaranteed computation offloading for mobile-edge cloud computing. IEEE Wireless Communications Letters. 2017;**6**(6):774-777

[6] Chen S, Chen Z, Gu S, Chen B, Xie J, Guo D. Load balance awared data sharing systems in heterogeneous edge environment. In: 2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS). Hong Kong: IEEE; 2020. pp. 132-139

[7] Han D, Chen W, Fang Y. A dynamic pricing strategy for vehicle assisted mobile edge computing systems. IEEE Wireless Communications Letters. 2018;**8**(2):420-423

[8] Xue J, Guan X. Collaborative computation offloading and resource allocation based on dynamic pricing in mobile edge computing. Computer Communications. 2023;**198**:52-62

[9] Siew M, Cai D, Li L, Quek TQ. Dynamic pricing for resource-quota sharing in multi-access edge computing. IEEE Transactions on Network Science and Engineering. 2020;**7**(4):2901-2912

[10] Wang X, Duan L. Dynamic pricing and capacity allocation of UAV-provided mobile services. In: Proceedings of IEEE Conference on Computer Communications. Paris, France: IEEE; 2019. pp. 1855-1863

[11] Chen S, Chen B, Tao X, Xie X, Li K. An online dynamic pricing framework for resource allocation in edge computing. Journal of Systems Architecture. 2022;**133**:102759. Available from: https://www.sciencedirect.com/science/article/pii/S1383762122002442

[12] Chen S, Zhang Q, Dong X, Tao X, Li K, Qiu T, et al. Sublessor: A cost-saving internet transit mechanism for cooperative MEC providers in industrial internet of things. IEEE Transactions on Industrial Informatics. 2023;**19**(9):9855-9866

[13] Qiu H, Zhu K, Luong NC, Yi C, Niyato D, Kim DI. Applications of auction and mechanism design in edge computing: A survey. IEEE Transactions on Cognitive Communications and Networking. 2022;**8**(2):1034-1058

[14] Hung YH, Wang CY, Hwang RH. Combinatorial clock auction for live video streaming in mobile edge computing. In: Proceedings

of IEEE Conference on Computer Communications Workshops. Honolulu, HI, USA: IEEE; 2018. pp. 196-201

[15] Jiao Y, Wang P, Niyato D, Xiong Z. Social welfare maximization auction in edge computing resource allocation for mobile blockchain. In: Proceedings of IEEE International Conference on Communications. Kansas City, MO, USA: IEEE; 2018. pp. 1-6

[16] Bahreini T, Badri H, Grosu D. Mechanisms for resource allocation and pricing in Mobile edge computing systems. IEEE Transactions on Parallel and Distributed Systems. 2022;**33**(3):667-682

[17] Sun W, Liu J, Yue Y, Zhang H. Double auction-based resource allocation for mobile edge computing in industrial internet of things. IEEE Transactions on Industrial Informatics. 2018;**14**(10):4692-4701

[18] Wang Q, Guo S, Liu J, Pan C, Yang L. Profit maximization incentive mechanism for resource providers in mobile edge computing. IEEE Transactions on Services Computing. 2022;**15**(1):138-149

[19] Zhang Y, Wang CY, Wei HY. Parking reservation auction for parked vehicle assistance in vehicular fog computing. IEEE Transactions on Vehicular Technology. 2019;**68**(4):3126-3139

[20] Zhou H, Wu T, Chen X, He S, Guo D, Wu J. Reverse auction-based computation offloading and resource allocation in mobile cloud-edge computing. IEEE Transactions on Mobile Computing. 2023;**22**(10):6144-6159

[21] Xu Q, Su Z, Wang Y, Dai M. A trustworthy content caching and bandwidth allocation scheme with edge computing for smart campus. IEEE Access. 2018;**6**:63868-63879

[22] Tao X, Ota K, Dong M, Qi H, Li K. Cost as performance: VNF placement at the edge. IEEE Networking Letters. 2021;**3**(2):70-74

[23] Deng J, Cai H, Chen S, Ren J, Wang X. Improved flow awareness among edge nodes by learning-based sampling in software defined networks. Mobile Networks and Applications. 2022;**27**:1867-1879

[24] Chen S, Chen B, Xie J, Liu X, Guo D, Li K. Joint service placement for maximizing the social welfare in edge federation. In: 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS). Tokyo, Japan: IEEE; 2021. pp. 1-6

[25] Li J, Guo D, Xie J, Chen S. Availability-aware provision of service function chains in mobile edge computing. ACM Transactions on Sensor Networks. 2023;**19**(3):1-28

[26] Chen S, Li J, Chen B, Guo D, Li K. vHSFC: Generic and agile verification of service function chain with parallel VNFs. In: 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Rio de Janeiro, Brazil: IEEE; 2023. pp. 498-503

Chapter 7

Understanding Citizen Engagement in the Era of Smart Cities

Stella Bvuma

Abstract

Beyond a simple technological demonstration, widespread public participation is necessary for the development of smart cities. To guarantee that underrepresented perspectives are heard, this entails tackling disenfranchisement, fostering digital literacy, and utilising a variety of engagement strategies (online forums and community gatherings). Long-term participation builds trust and gives communities the ability to co-design audit procedures and solutions through advisory committees and participatory budgeting. Citizen engagement in the era of smart cities is necessary for developing effective governance strategies, community collaboration and ensuring that technological advancement addresses the diverse needs of the population.

Keywords: collaborative cities, data-driven governance, digital literacy, long-term involvement, civic engagement, inclusion, marginalised communities, co-design

1. Introduction

The smart city movement aims to integrate digital technology and data analytics into city operations and infrastructure, resulting in more sustainable, liveable, and efficient urban settings. However, achieving this goal would necessitate placing people first and actively involving the public in the process of creating better cities. Cities are for people, after all, and technology should support human goals and wants [1]. Still, a lot of smart city projects struggle to achieve the elusive aim of meaningful community engagement. Promoting inclusive involvement that extends beyond the "usual suspects" of active community groups is a challenge for governments. Those on the margins frequently lack the time, money, internet connection, or technological know-how to share their viewpoints. Cultural and linguistic obstacles prevent minorities and immigrants from participating in public life. Younger people of working age find it difficult to participate because of their juggling of work and family obligations [2]. As a result, plans for smart cities often neglect to address genuine citizen issues, which erode public confidence in hierarchical, bureaucratic planning. However, there are instances of cities using innovative involvement methods to mould smart policies centred on people hackathons that capitalise on young creativity, and participatory budgeting campaigns are just two examples. The development of digital platforms has made it easier to include previously marginalised voices in procedures such as crowdsourcing on smartphones, localised social networks, interactive planning tools, and public consultations.

Nevertheless, underlying socioeconomic injustices that impede participation cannot be resolved by technology alone [3]. Whether via digital literacy initiatives, community organising at the grassroots level, or granting access to places for participation, capacity building is as important. In the end, cities need to establish long-term involvement through the establishment of permanent citizen advisory committees that work with organisations on the creation of smart cities. Achieving inclusion well pays off. By focusing on locals, smart cities may provide economic opportunities in underprivileged areas by offering activities such as small business incubators and skill development programmes. By shedding light on inequity, new statistics help vulnerable groups receive better public services. Community knowledge also ensures that technology implementation is grounded in real requirements, including those related to environment, health, safety, and transit. The smart city revolution can deliver on its promise of better futures for all urban residents if people oversee driving change [4].

2. Smart cities: a new look at urban management

The concept of smart cities makes use of real-time data, cognitive analytics, and connected devices to give metropolitan areas the knowledge they require for improved infrastructure performance, more effective resource management, and creative public services catering to the requirements of citizens. Significant gains in quality of life are anticipated from data-driven solutions, which include less traffic jams because of adaptive signals, more dependable public transport thanks to live monitoring applications, reduced community energy usage according to smart grid technology, and early notification of pipe failures that stop water main breaks. However, without significant input and engagement from locals, even the most advanced models and algorithms can ensure successful smart cities [5]. Participation by citizens is still essential at every level, from selecting which technologies to use first to offering critical localised knowledge that anchors digital tools in reality on the ground. For example, in the absence of hyper-local viewpoints, centralised efforts to forecast ideal routes for ridesharing services could miss typical pickup locations or underestimate travel times through certain neighbourhoods. Analogously, digitising requests for municipal services, such as patching potholes, will not go far if platforms overlook citizen technological competency limitations that prevent uptake in underprivileged areas. Thus, rather than adding ongoing citizen participation as an afterthought when projects encounter implementation roadblocks, municipal managers should prioritise this in smart city roadmaps. Innovative approaches to citizen engagement include crowdsourcing hyper-local data points through smartphone applications to enhance decision-making and localised social networks for peer debate of neighbourhood concerns. Of course, addressing exclusionary hurdles related to language, digital access, mobility limitations, and the historical disenfranchisement of marginalised urban communities is necessary for true public engagement. When inclusion is done well, smart cities may use data-driven innovation to advance equity by leveraging technology to help underprivileged areas rather than letting it exacerbate existing inequalities. Smart cities may realise their revolutionary potential and benefit all citizens when people oversee improving infrastructure and services [6].

3. Enhancing citizen engagement through edge computing

It is essential to demonstrate how Edge Computing technologies facilitate data-driven decision-making and solve real-world urban problems in order to integrate their role in improving smart city citizen participation. With the help of edge computing, data can be processed much closer to its point of generation, thus cutting down on latency. Applications like public safety and traffic management rely on real-time data processing; therefore, this is essential for making smart city services more responsive to citizens' requirements [7]. Edge Computing improves data security and privacy by processing data locally at the edge instead of sending it back to a central cloud. When people provide their own personal information for participatory applications, this becomes quite crucial [8]. Smart cities can be more responsive and adaptive with the help of edge computing, which facilitates decentralised decision-making processes. Citizen engagement in urban management is made more effective by local processing, which enables faster reactions to changes in urban environments [9]. Smart cities can handle the massive amounts of data produced by urban Internet of Things (IoT) devices with the help of scalable computing resources provided by edge computing, which prevents central computing facilities from being overwhelmed. In fast-paced urban areas, this scalability is essential for effective resource management [10]. In smart cities, where IoT devices are becoming more common, edge computing is useful for controlling the flow of data and meeting the processing demands of these devices. Citizens rely on these for a wide range of services, including public transit, healthcare, and environmental monitoring, all accessible through smartphone applications [11]. These ideas will be incorporated into the chapter to give a thorough review of how Edge Computing technologies improve data management, security, responsiveness, and scalability, which in turn boost citizen participation in smart cities.

Integrating edge computing into smart city frameworks offers a multifaceted approach to improving urban management and enhancing citizen engagement through localised data processing. Edge computing facilitates real-time data processing by managing information close to its source, which is crucial for applications like traffic management and public safety, where immediate response is essential. This capability not only improves urban living experiences but also supports critical responses that can significantly impact safety and efficiency [12]. Moreover, edge computing enhances data security and privacy by processing data locally, reducing the risk of breaches and increasing public trust in smart city initiatives. This aspect of edge computing is vital for maintaining the integrity and confidentiality of citizen data, a fundamental component in fostering trust and encouraging broader participation in smart city programmes [13]. The decentralised decision-making enabled by edge computing allows cities to be more adaptive and responsive to changes and needs at the local level. This approach not only supports the scalability necessary to manage the vast data generated by urban Internet of Things (IoT) devices but also ensures that city operations can efficiently scale without overloading central computing resources. This scalability is crucial for the sustainable operation of smart cities [4]. Lastly, edge computing empowers citizens by enabling real-time, location-based interaction with smart city platforms, thereby enhancing civic participation. This interaction helps ensure that smart city initiatives are not only technologically advanced but also grounded in the needs

and participation of the local population, fostering an inclusive urban environment [14]. Through these mechanisms, edge computing significantly contributes to the efficiency and inclusivity of smart cities, ensuring that technology serves the populace effectively and that urban environments are both progressive and responsive to their inhabitants.

4. Changing citizen engagement paradigms

The smart city revolution demands ongoing public participation through interactive platforms and localised data streams, going beyond sporadic town halls and polls. Cities require resident viewpoints to contextualise and make sense of emergent patterns as sensors, metres, and IoT devices generate real-time information on urban dynamics. Additionally, unprecedented citizen co-creation of new services linked with grassroots goals is made possible by mobility applications and digitalised municipal operations. Universal access also supports the involvement of marginalised voices in issue diagnosis and solution development on a larger scale. More cooperative, nuanced public-private partnerships in the pursuit of common aims and the common good are made possible by the sheer number and variety of participation options [6].

4.1 From hierarchical to collaborative

In order to maximise the benefits of local knowledge, progressive, smart cities aim to involve citizens directly in the planning, prioritisation, and solution-creation processes, as opposed to just informing them of existing plans. For example, historically underprivileged neighbourhoods might identify chronic infrastructure deficiencies that city models ignored using an app that crowdsources localised flood data, indicating the areas most in need of stormwater system repairs [15]. Real-world complications are included in simulation systems that guide transportation development or public Wi-Fi growth, with individuals offering on-the-ground opinions. Residents become empowered co-creators committed to ensuring collaborative solutions robustly react to community needs, rather than passive consumers of government services, by actively participating in designing smart city programmes that are still in the planning stages [16].

4.2 Moving from in-person involvement to digital inclusion

Smart cities enable more citizens to influence local objectives and development in ways that suit their schedules and communication preferences by augmenting in-person community meetings with digital engagement tools. Citizens with mobility impairments or unable to attend public hearings during business hours can nonetheless contribute using web portals and mobile apps. Options for offline and multilingual involvement assist in removing obstacles for non-English speaking families and those without access to the internet at home. In addition, online forums allow for more dynamic, peer-to-peer discussions about important problems rather than just official statements being disseminated in one direction. In order to create truly inclusive digital citizen involvement, it is ultimately necessary to prioritise developing the capacities of marginalised groups in addition to cutting-edge technology. This will ensure that innovative smart city solutions are developed from the bottom up rather than merely from the top down [4].

4.3 Citizen input based on data

Through the use of social media listening tools, sensor networks, and crowd-sourced data streams, leaders of smart cities may have comprehensive insight into the concerns and objectives of citizens about a more flexible and data-driven government. Analytics showing an increase in pedestrian injuries and noise complaints on a busy route, for instance, may prompt officials to quickly remodel the corridor with traffic-calming features before tensions worsen. Digitally improved citizen feedback loops enable administrators to identify problems early and make necessary corrections, making responsiveness the new necessity. This builds public trust in government by promoting dependability and openness [4].

5. Opportunities and difficulties

Smart cities confront challenges in ensuring engagement projects live up to their potential, despite the abundance of options to include citizens in data-enhanced municipal administration. Continuous digital gaps run the risk of removing opinions from underrepresented groups who are ill-equipped to use technology. Limiting involvement diversity may result from favouring some forms of participation over inclusive grassroots organising. A lack of institutional processes for integrating opinions into decision-making may encourage mistrust or weariness from consultation. Smart cities cannot ensure that technology-enabled civic involvement will increase equitable responsiveness, accountability, or transparency if structural hurdles are not addressed head-on [6].

5.1 The gap in digital

While smart cities promise more transparent and responsive governance, these benefits hinge on comprehensive public digital inclusion. However, socioeconomic barriers like poverty, language gaps, and disability continue blocking many marginalised urban groups from accessing engagement platforms and using tech tools optimally. For instance, lower-income neighbourhoods may lack affordable broadband or public computers to provide inputs on the location of new data-driven sustainability infrastructure. Without addressing unequal access through targeted capacity building and access provisions, vulnerable residents risk being left out of data-enriched civic participation loops that determine their communities' futures [17].

5.2 Concerns about privacy

Establishing strong public confidence in responsible data management is necessary to realise the collaborative governance potential of smart cities. On the other hand, worries about data breaches, privacy invasions and monitoring, or opaque automated decision-making might deter participation. For example, citizens' concerns about privacy may prevent them from giving real-time ridesharing firms the data they need to improve transportation planning unless strict cyber security and moral usage policies are first established. By reducing the possibility of misuse through localised data storage, encryption, anonymity, and open algorithm auditing, information exchange may be made more secure and facilitate mutually beneficial public-private data partnerships for the benefit of society as a whole [6].

5.3 Accountability and openness

Once early consultations are over, smart cities cannot assume that public engagement is taken for granted without consistently proving that citizen comments directly impact policies. Delivering on the promises of collaborative governance requires keeping lines of communication available for clear, two-way communication. Examples of these channels may include interactive web dashboards that show how proposals are put into practice or multilingual citizen working groups that track developments. Similarly, easy-to-use feedback loops that let locals voice their concerns, along with prompt city issue status updates even for items that are delayed, help create a long-lasting trust that promotes continued involvement. Communities that are engaged are rewarded when cities invest in ongoing communication [18].

5.4 Despite these difficulties, there are a tonne of potential

Although there are many obstacles in the way of achieving widespread civic engagement in smart cities, doing so might have a huge positive impact on urban inhabitants' quality of life. By means of responsive policies that target localised objectives, inclusive involvement around data-driven decision-making has the potential to elevate marginalised communities. From the ground up, responsible and representative artificial intelligence (AI) governance frameworks are maintained by active public scrutiny and collective intelligence. Input from empowered individuals also enables cities to resiliently and dynamically modify infrastructure and service delivery to address new issues. When civic engagement is fully utilised to its fullest extent, inclusive technology access and digital capabilities serve as the cornerstone of progressive, fair, and equitable smart cities that prioritise the welfare of all [19].

5.5 Enabled citizens

Smart cities that prioritise participatory government allow for continuous public input and community objectives to guide development, as opposed to restricting inhabitants to sporadic elections. Applications (Apps) that crowdsource neighbourhood infrastructure requirements or digitally enabled participatory budgeting procedures, for instance, establish direct democracy methods that give residents more control over local decision-making. Active participation encourages more involved and civic-minded groups that are committed to jointly defending their rights and creating settings that are responsive to the goals of the grassroots rather than merely following directives from above. Empowered participation is the engine that propels fair growth in smart cities centred on democratic values [20].

5.6 Making well-informed decisions

When decision-makers have access to real-time feedback from citizens in smart cities, they can create more effective and responsive policies to solve issues that arise locally. For example, citizen-generated data identifying transportation bottlenecks or neighbourhoods that flood frequently offers critical micro-targeting information about where infrastructure investments should be prioritised. Additionally, consistent data flows surrounding new issues allow for more flexible governance by directing decisions on everything from the growth of public Wi-Fi to hotspot pollution reduction. Cities fill up knowledge gaps that enable needs-based, inclusive

planning by utilising data derived directly from the lived experiences of various urbanites [21].

5.7 Enhanced provision of services

By using digital channels to provide continuous resident feedback and engagement, smart cities may unlock enormous potential for not just resolving specific infrastructure problems but also proactively co-designing new services that improve communities. For instance, applications that solicit feedback from users on issues they have with public transport or ideas for better routes or timetables enable organisations

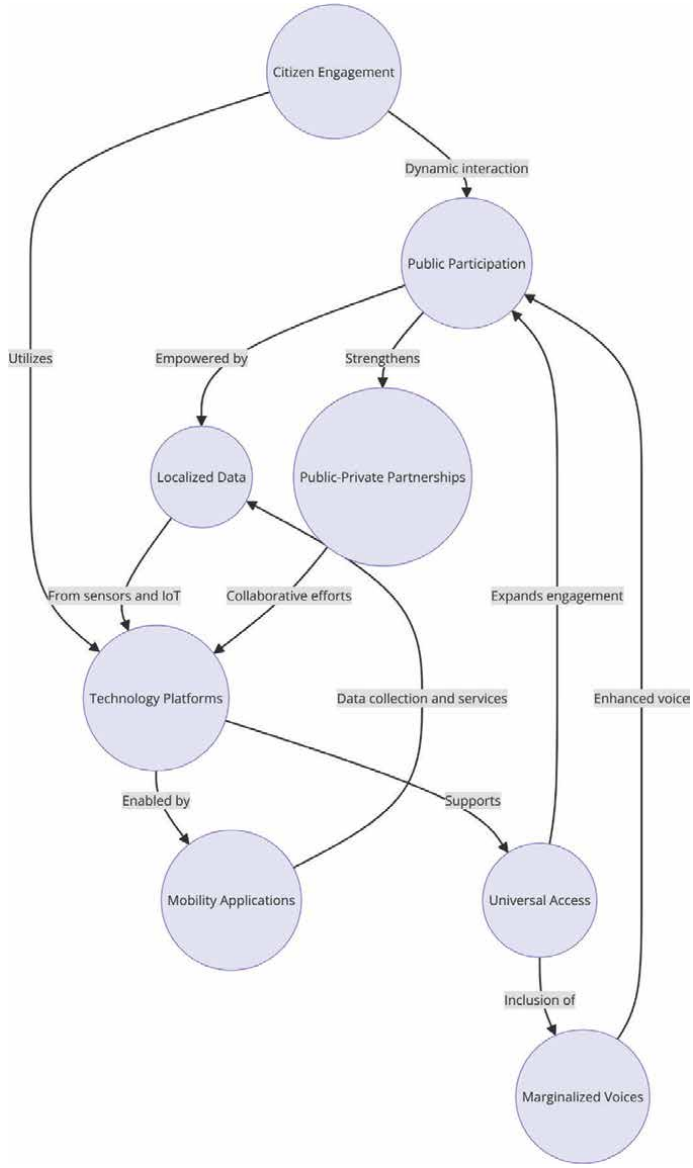

Figure 1.
Interconnected framework of citizen engagement in smart cities.

to work with consumers to improve accessibility and service quality. Through focused enhancements aimed at addressing experience gaps and near real-time response to end-user demands across municipal domains, cities fulfil their responsibility to profoundly improve the daily lives of their citizens while simultaneously increasing approval and usage of public services [22].

The graph diagram (**Figure 1**) illustrates the intricate interconnections within the smart city framework, with a specific emphasis on citizen engagement and the incorporation of digital platforms to augment public participation and joint endeavours. The nodes symbolise crucial components such as citizen involvement, public involvement, technological platforms, localised data, mobility applications, equal access, underrepresented voices, and collaborations between the public and commercial sectors. The boundaries illustrate the ever-changing connections and impacts among various components, demonstrating their individual contributions to the data-based decision-making process and the engagement of the entire community. There is a dire need to emphasise the robustness of our scientific approach and the significance of these findings in addressing current urban concerns.

6. Metrics for measuring engagement success

Recent research has identified many metrics and methodologies that can be used to assess the effectiveness of citizen participation and its influence on the outcomes of urban planning. Evaluating the degree of involvement and encompassing the variety and number of participants engaged (in terms of gender, age, and stakeholder category) is of utmost importance. Metrics should encompass both the scope (number of participants) and intensity (level of engagement) of involvement [23]. Digital platforms, town halls, and surveys are all examples of engagement tools; it is important to assess their reach and variety as well as the efficacy of communication tactics. To do this, we must examine how often, how clear, and how easy it is to obtain the public domain data [24]. Metrics should take into account the extent to which programmes educate the public and give them the tools they need to participate actively in urban planning. Programmes and materials for instruction that increase familiarity with digital tools and urban planning are part of this category [25]. It is important to measure the real impact of public feedback on city planning, including how it is used to make decisions and how much it changes the end result. Modifications to plans can be implemented in response to citizen comments and to meet community requirements [26]. It is critical to assess how long engagement initiatives will last. As a transition from episodic to continuous engagement occurs, metrics should track how often citizens participate and whether permanent mechanisms or venues are put in place to continue soliciting citizen feedback over time [27]. A more thorough and useful guide for evaluating and assessing the impact of citizen engagement in smart city programmes can be created by including these particular measurements and frameworks.

7. Best practices and case studies

Smart city citizen engagement errors are still prevalent, yet there are groundbreaking instances that demonstrate the transformational power of inclusion when implemented properly. Diverse creative methods, such as crowdsourced public transportation scheduling or artificial intelligence (AI) chatbots developed in collaboration

with marginalised youth to lead housing aid applications, offer strategies to ensure that technology-enabled urban solutions are centred on community empowerment. Analysing these accomplishments provides guidance and motivation for implementing smart cities that are participatory and egalitarian.

7.1 The Decidim platform in Barcelona

The Decidim site in Barcelona serves as an example of how digital platforms may support extensive grassroots public discussions that have a genuine impact on budgetary allocations and policy. Residents may vote on proposals ranging from public housing to mobility enhancements, discuss priorities with neighbours, and submit suggestions utilising the flexible e-participation system. The outcomes directly influence significant city planning and finance decisions. Decidim is an example of participatory democracy in action; it ensures public supervision from ideation to implementation by providing continuous, transparent feedback loops and accountability monitoring. This keeps citizens involved as projects go from ideation to finalisation [28].

7.2 Open innovation square in Seoul

Open Innovation Square in Seoul offers the public a physical area where they may work together to develop and evaluate possible smart city solutions for regional urban problems. The hub, which brings together academic institutions, start-ups, government agencies, and regular people to discuss open data and collaborative ideation, has promoted creative, grassroots pilot projects like applications to track parking availability in business areas or sensors to identify flooded pedestrian underpasses that require drainage repairs after monsoons. The Square ignites public buy-in and ground-trusting for human-centric smart urban change by empowering communities as collaborators rather than just recipients of top-down innovations [29].

7.3 The city of things network in Amsterdam

The City of Things project in Amsterdam uses Internet of Things technology to promote broad community engagement with urban data in order to promote sustainability. By utilising sensor-equipped public installations such as Wi-Fi-enabled smart benches, locals may actively contribute to the crowdsourcing of environmental knowledge on issues like traffic, noise, air quality, and energy use. The initiative also uses surveys and open forums to get public input on how to strike a balance between data protection and innovative prospects. The effort offers opportunities for positive and trust-cantered public-private smart city data cooperation by blending digital and physical interaction approaches [30].

8. Empowering communities and improving public services

Bvuma and Bwalya [31] examine the idea of utilising open data to improve public services and empower communities, with a particular emphasis on local government in South Africa. The writers explore governance strategies that help open data projects generate value for the general population [32]. The concepts put out in their study highlight how, in the South African context, open data may have a beneficial influence

on community empowerment, local government, and the standard of public services. Their insight into their conclusions are also highlighted in an article, namely: "*Governance Models for Creating Public Value in Open Data Initiatives*" [33].

9. Tech challenges: unseen impacts in smart cities

There are a number of important topics that should be prioritised, according to the research: There are serious concerns about the security and privacy of smart city technologies that gather and handle massive volumes of personal data. Preventing breaches that could disclose sensitive information requires careful and ethical handling of this data [34]. The prospect of ongoing surveillance in smart cities gives rise to apprehensions regarding the erosion of privacy and civil liberties. It is imperative to consider the potential misuse of these technologies by authorities or other entities, which could result in the establishment of a surveillance state [27]. Advanced technological solutions in metropolitan areas have the potential to unintentionally exacerbate social disparities by primarily benefiting affluent communities while disregarding or excluding lower-income groups. Smart city projects should prioritise inclusive designs that cater to the requirements of all inhabitants, irrespective of their socioeconomic background [35].

People who aren't comfortable with technology run the danger of being left out as cities rely more and more on complicated systems. A "digital divide" results when certain demographics, such as the elderly, those with disabilities, and others, are unable to use digital technologies [36]. Smart cities must address the ethical ramifications of technology utilisation, namely with data ownership, permission, and the openness of algorithms that impact public services and individual lives. Legal frameworks must adapt to safeguard citizens against potential misconduct and enable responsible utilisation of smart city technologies [17].

This concept map (**Figure 2**) visually organises the key components and outcomes of smart city frameworks. It emphasises the role of technologies like connected devices and real-time data in improving public services and urban management, while also highlighting the importance of addressing challenges such as technological competency and exclusionary hurdles to enhance overall city life and equity.

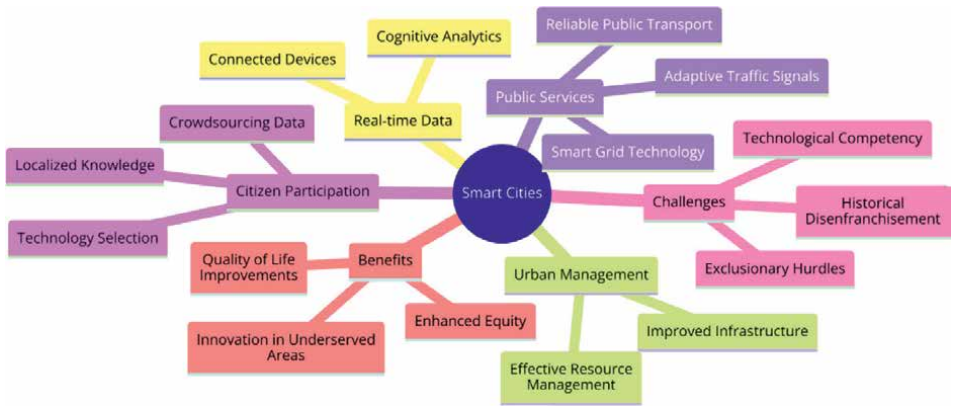

Figure 2.
Interconnected elements of smart city development.

10. Recommendations

Within the quickly changing field of urban development, the emergence of smart cities has ushered in a new age characterised by technology innovations meant to improve inhabitants' quality of life [14]. The idea of citizen involvement, or the dynamic and reciprocal connection between citizens and municipal authorities, is essential to the development of smart cities. In order to promote a more profound comprehension of citizen participation in the smart city era, a number of suggestions surface as essential for the successful execution and maintenance of this creative urban ecosystems. The first and most important step is to create strong and intuitive digital platforms. Smart cities use technology to engage with its residents, and having user-friendly Internet and mobile application interfaces may help with smooth communication. Information regarding municipal services, events, and projects should be easily accessible to the public so they may keep informed and involved in community civic life. In order to guarantee inclusion, it is also crucial to promote digital literacy. As smart cities develop, it is critical to close the digital divide by offering tools and training that enable people from all walks of life to use and understand digital platforms. Smart cities may guarantee that citizen involvement is not restricted to a tech-savvy few but is available and advantageous to all citizens by promoting digital literacy. Furthermore, encouraging openness is essential to fostering confidence between the public and local government. Open data projects may give the public information on city finances, operations, and decision-making procedures. In addition to fostering confidence, transparency encourages residents to take an active role in determining their city's destiny. It is possible to use interactive dashboards and real-time data sharing to inform the public about ongoing projects, environmental data, and urban planning techniques. Additionally, active participation in decision-making processes should be an element of public participation, going beyond simple information consumption. The use of public consultation and feedback methods, such as town hall meetings, surveys, and participatory budgeting, confers authority to residents to express their viewpoints and participate in the development of policies and initiatives. This inclusive strategy encourages a sense of ownership and belonging among inhabitants while strengthening the social fabric. Lastly, in order to allay worries about data abuse and spying, privacy and security must come first. Large-scale data gathering and analysis are keys to smart cities. Thus, it is essential to put strong data protection mechanisms in place to guarantee that citizen data is managed in an ethical and responsible manner. Establishing precise rules and regulations for the use and storage of data can promote confidence and allay privacy worries.

11. Conclusion

If smart cities are to enhance lives instead of merely deploy flashy technology, and then true public involvement is much more than just a box-checking exercise. Even though civic engagement presents many complicated issues, inclusive public involvement serves as a vital foundation for achieving smart urban administration that is trustworthy, responsible, and responsive. Sustained community feedback loops fill vital information gaps when top-down policies and algorithms fail to capture hyper-local reality or react in time to looming crises. To address specific pain points, neighbourhood assemblies, internet polling platforms, and crowdsourced data flows about malfunctioning infrastructure or delayed services highlight issues on the ground.

Furthermore, addressing past disenfranchisement and arranging priorities around the needs of vulnerable communities are achieved by providing excluded and marginalised groups with a continuous voice in codesigning solutions. Of course, addressing unequal access to technology and enhancing digital competencies in underprivileged areas via computer centres and training programmes are necessary for true participation. Encouraging participation in many languages and providing physical, non-digital routes for members who are not tech-savvy or who have impairments are also necessary for holistic inclusion. Furthermore, maintaining civic confidence and interest requires long-term engagement continuity as opposed to occasional, brief drives. Alongside digital town halls and participatory budgeting efforts, long-term community advisory committees that have the authority to audit algorithms, oversee data sharing agreements, and openly examine progress transparency become essential. By allowing individuals to continuously guide the process, smart cities have the potential to greatly enhance collaboration. Municipal leaders may transform smart city initiatives from exclusionary technocratic agendas to really empowering bottom-up platforms elevating the marginalised by emphasising inclusive involvement. Now is the moment to take on the most formidable obstacles to participation and to create the kind of constructive disruption that will make civic engagement the backbone of data-driven urban change.

12. Further reflections

There is a need to emphasise the robustness of our scientific approach and the significance of this chapter in addressing current urban concerns. The chapter incorporated the significance of addressing the socioeconomic ramifications of smart city technologies. This chapter took into consideration the linguistic choice and ensured that the writing is accessible to a broad readership and prioritised, which is why there is a need to focus on using consistent and straightforward language. The author has ensured that intricate concepts are presented in a more accessible manner while maintaining the intellectual rigour of the research. As a result, this chapter serves as a helpful reference for both specialists in the area and the general public with an interest in the future of urban living.

Author details

Stella Bvuma
University of Johannesburg—SCiiS, Johannesburg, South Africa

*Address all correspondence to: stellab@uj.ac.za

References

[1] The people-centered smart city: Putting citizens at the heart of sustainable urban development by UN-habitat. Available from: https://cmte.ieee.org/futuredirections/tech-policy-ethics/nov-2021/best-practices-for-community-engagement-in-smart-cities/

[2] Reassessing the smart cities movement by smart cities dive. Available from: https://www.smartcitiesdive.com/

[3] The smart cities council. Available from: https://www.smartcitiescouncil.com/

[4] The people-centered smart city: Putting citizens at the heart of sustainable urban development by UN-habitat. Available from: https://unhabitat.org/programme/legacy/people-centered-smart-cities

[5] Smart cities for all: a guide to equitable and inclusive design by world resources institute. Available from: https://publications.wri.org/transformations-equitable-sustainable-cities

[6] The Rockefeller foundation: 100 resilient cities program. Available from: https://www.rockefellerfoundation.org/100-resilient-cities/

[7] Maltezos E, Karagiannidis L, Dadoukis A, Petousakis K, Misichroni F, Ouzounoglou E. Public safety in smart cities under the edge computing concept. In: IEEE International Mediterranean Conference on Communications and Networking. New York City, USA: IEEE; 2021. pp. 88-93

[8] Khan ZA, Abbasi AG, Pervez ZJC, Practice C. Blockchain and edge computing–based architecture for participatory smart city applications. Concurrency and Computation Practice and Experience. 2019;**32**:1-20

[9] Jararweh Y, Otoum S, Ridhawi IA. Trustworthy and sustainable smart city services at the edge. Sustainable Cities and Society. 2020;**62**:102394

[10] Khan LU, Yaqoob I, Tran NH, Kazmi SMA, Tri ND, Hong CS. Edge-computing-enabled smart cities: A comprehensive survey. IEEE Internet of Things Journal. 2019;**7**:10200-10232

[11] Liu Q, Gu J, Yang J, Li Y, Sha D, Xu M, et al. Cloud, Edge, and Mobile Computing for Smart Cities. Urban Informatics. The Urban Book Series. Singapore: Springer; 2021. pp. 757-795

[12] Rethinking the smart city: Putting people first by MIT technology review. Available from: https://www.technologyreview.com/2021/04/28/1023104/smart-cities-urban-technology-pandemic-covid/

[13] Rethinking the smart city: Putting people first by MIT technology review. Available from: https://quizlet.com/45050993/personal-finance-chapter-16-flash-cards/

[14] People-centered smart cities by UN-habitat. Available from: https://unhabitat.org/people-centered-smart-cities

[15] Smart cities for all: A guide to equitable and inclusive design by world resources institute. Available from: https://g3ict.org/publication/smart-cities-for-all-a-vision-for-an-inclusive-accessible-urban-future

[16] A guide to equitable and inclusive design by world resources institute.

Available from: https://www.wri.org/research/7-transformations-more-equitable-sustainable-cities

[17] Seven transformations for more equitable and sustainable cities by world resources report. Available from: https://www.wri.org/research/7-transformations-more-equitable-sustainable-cities

[18] Granier B, Kudo H. How are citizens involved in smart cities? Analysing citizen participation in Japanese "Smart Communities". Information Polity. 2016;**21**(1):61-76

[19] Belausteguigoitia J, Alonso I, Chueca A, Elizegi A, Hierro S, Olavarri L, et al. Measuring Participation: A Comparative Study of Citizen Engagement Processes in Urban Planning. Southampton, UK: WIT Press; 2021

[20] Cortés-Cediel ME, Cantador I, Bolívar M. Analyzing citizen participation and engagement in European smart cities. Social Science Computer Review. 2019;**39**:592-626

[21] Garg S, Mittal S, Sharma S. Role of e-trainings in building smart cities. In: 8th International Conference on Advances in Information Technology. Vol. 111. 2017. pp. 24-30

[22] Yonggang H. Impact of communication quality in facilitating citizen participation in urban planning. Advances in Intelligent Systems and Technologies. 2022:33-43

[23] Empel CV. The Effectiveness of Community Participation in Planning and Urban Development. Southampton, UK: WIT Press; 2008

[24] Przeybilovicz E, Cunha MA, Geertman S, Leleux C, Michels A, Tomor Z, et al. Citizen participation in the smart city: Findings from an international comparative study. Local Government Studies. 2022;**48**(1):23-47

[25] Barcelona pact for the safeguarding of human rights in the city. Available from: https://www.decidim.barcelona/?locale=es

[26] ICO 2023 SEOUL. Available from: http://icomeeting.org/

[27] Bvuma S, Joseph BK. Empowering communities and improving public services through open data: South African local government perspective. In: Rodríguez Bolívar M, Bwalya K, Reddick C, editors. Governance Models for Creating Public Value in Open Data Initiatives. Public Administration and Information Technology, Vol. 31. Cham: Springer; 2019. DOI: 10.1007/978-3-030-14446-3_7

[28] Barandiaran XE, Calleja-López A, Monterde A, Romero C. Decidim: A brief overview. In: Decidim, A Technopolitical Network for Participatory Democracy. SpringerBriefs in Political Science. Cham: Springer; 2024. pp. 1-33

[29] Braun T, Fung BCM, Iqbal F, Shah B. Security and privacy challenges in smart cities. Sustainable Cities and Society. 2018;**39**:499-507

[30] Elmaghraby AS, Losavio MM. Cyber security challenges in smart cities: Safety, security and privacy. Journal of Advanced Research. 2014;**5**:491-497

[31] Bvuma S, Bwalya KJ. Fog computing in a developing world context: Jumping on the bandwagon. In: Mahmood Z, editor. Fog Computing. Cham: Springer; 2018. DOI: 10.1007/978-3-319-94890-4_4

[32] Eckhoff D, Wagner IJ. Privacy in the smart city—Applications, technologies,

challenges, and solutions. IEEE Communications Surveys & Tutorials. 2018;**20**:489-516

[33] Cui L, Xie G, Qu Y, Gao L, Yang Y. Security and privacy in smart cities: Challenges and opportunities. IEEE Access. 2018;**6**:46134-46145

[34] Zhang K, Ni J, Yang K, Liang X, Ren J, Shen XJ. Security and privacy in smart city applications: Challenges and solutions. IEEE Communications Magazine. 2017;**55**:122-129

[35] Michael EM. Lucidity and science. Interdisciplinary Science Reviews. 1997;**1**(22):N0-N3

[36] Bibri SE, Krogstie J. On the social shaping dimensions of smart sustainable cities: A study in science, technology, and society. Sustainable Cities and Society. 2017;**29**:219-246

Edited by Yu Chen and Ronghua Xu

Edge computing technology is the cornerstone of digital advancement as smarter cities and unified industries continue to rise. The book aims to provide a deep dive into state-of-the-art developments and practical uses of edge computing in the current technological scenario. The first part is focused on new architectures and presents a proposal for Autonomous and resilient Edge (AR-Edge) computing, which combines AI, Software-Defined Networks (SDN), and blockchain technologies to generate transparent networks in smart cities. It investigates how Edge Computing can augment AI applications with lower latency and computational costs. It also introduces a green computation paradigm for Internet of Medical Things (IoMT) devices that would optimize outcomes in critical areas such as robotic surgery or autonomous vehicles. The second section of the book looks at a range of application scenarios. Chapter 4 presents in-depth privacy and security features as well as real-time monitoring frameworks for the safety of smart homes using IoT to improve their facilities. Then, it examines the implications of 5G-based edge-cloud capabilities for industrial energy facilities to illustrate how they contribute toward efficiency and innovation in Industry 4.0. It also covers the strategic deployment models and dynamic pricing strategies in edge computing services, enabling system performance and economic benefits. In conclusion, it underscores the need for citizen engagement in smart cities. It opines for inclusive participation by sharing digital literacy and varied ways to build equitable solution-based urban development. The transformative power of edge computing comes from the experts who are making it happen. This book summarizes the state-of-the-art developments and future research challenges for researchers, practitioners, and policymakers on edge computing, along with major trends that need to be considered. Take a ride with us through the age of edge computing and look at all the different ways we could make our lives in cities better.

Published in London, UK

IntechOpen

ISBN 978-0-85466-771-0