Wireless Networks

Jie Gao
Mushu Li
Weihua Zhuang

Connectivity and Edge Computing in IoT: Customized Designs and AI-based Solutions

Springer

Wireless Networks

Series Editor

Xuemin Sherman Shen, University of Waterloo, Waterloo, ON, Canada

The purpose of Springer's Wireless Networks book series is to establish the state of the art and set the course for future research and development in wireless communication networks. The scope of this series includes not only all aspects of wireless networks (including cellular networks, WiFi, sensor networks, and vehicular networks), but related areas such as cloud computing and big data. The series serves as a central source of references for wireless networks research and development. It aims to publish thorough and cohesive overviews on specific topics in wireless networks, as well as works that are larger in scope than survey articles and that contain more detailed background information. The series also provides coverage of advanced and timely topics worthy of monographs, contributed volumes, textbooks and handbooks.

** Indexing: Wireless Networks is indexed in EBSCO databases and DPLB **

More information about this series at http://www.springer.com/series/14180

Jie Gao • Mushu Li • Weihua Zhuang

Connectivity and Edge Computing in IoT: Customized Designs and AI-based Solutions

Springer

Jie Gao
Department of Electrical and Computer Engineering
Marquette University
Milwaukee, WI, USA

Mushu Li
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, ON, Canada

Weihua Zhuang
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, ON, Canada

ISSN 2366-1186 ISSN 2366-1445 (electronic)
Wireless Networks
ISBN 978-3-030-88742-1 ISBN 978-3-030-88743-8 (eBook)
https://doi.org/10.1007/978-3-030-88743-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The Internet of Things (IoT) is revolutionizing the world and impacting the daily lives of billions of people. Supporting use cases for households, manufacturers, transportation, agriculture, healthcare, and much more, IoT carries many potentials and expectations for prospering human society. Technologically, we are at an early stage of IoT development, aiming at connecting tens of billions of devices to make homes, communities, factories, farms, and everywhere else smart and automated. Tremendous efforts are necessary to advance IoT research and development.

Two cornerstones of IoT are data collection/exchange and data analysis. The former demands connectivity solutions, while the latter requires computing solutions. Due to the broad scope of IoT and the drastically different characteristics and requirements of IoT use cases, no "one-size-fits-all" design can meet the expectations of all use cases. Therefore, customizing connectivity or computing solutions for specific use cases is challenging yet essential. There are many system features and performance measures to consider in the customization, such as connection link density, resource overhead, transmission and computation delay, service reliability, energy efficiency, and device mobility, and making proper trade-offs among them is critical.

Accounting for all performance metrics and making optimal trade-offs can yield high complexity. Correspondingly, artificial intelligence (AI) solutions, such as neural networks and reinforcement learning, can become useful. Powered by AI methods, connectivity or computing solutions can learn from experience to handle the complexity, assuming that sufficient data are available for training. Specifically, AI can play various roles in IoT, including data traffic load prediction, access control, and computation task scheduling, to name a few.

In this book, we focus on connectivity and edge computing in IoT and present our designs for four representative IoT use cases, i.e., smart factory, rural IoT, Internet of vehicles, and mobile virtual reality. We thoroughly review the existing research in this field, including many works published in recent years. Then, through innovative designs, we demonstrate the necessity and potential of customizing solutions based on the use cases. In addition, we exploit AI methods to empower our solutions. The four research works included in this book serve a collective objective:

enabling on-demand data collection and/or analysis for IoT use cases, especially in resource-limited IoT systems. We hope that this book will inspire further research on connectivity and edge computing in the field of IoT.

Milwaukee, WI, USA Jie Gao

Waterloo, ON, Canada Mushu Li

Waterloo, ON, Canada Weihua Zhuang

July 2021

Acknowledgements

The authors would like to thank Professor Xuemin (Sherman) Shen at the University of Waterloo, the Editor of the *Wireless Networks* series, for his support in publishing this book and his comments and suggestions on its technical content.

We would also like to thank Professor Lian Zhao at the Ryerson University for her helpful discussions on the research presented in Chaps. 3 and 4.

In addition, we thank Dr. Xu Li, Professor Nan Cheng, and Conghao Zhou, who participated in the research discussed in Chaps. 2, 4, and 5, respectively.

We appreciate valuable discussions with the members of the Broadband Communications Research (BBCR) Lab at the University of Waterloo.

Special thanks to Susan Lagerstrom-Fife, Senior Editor at Springer New York, and Shina Harshavardhan, Project Coordinator for Springer Nature for their help during the preparation of this monograph.

Contents

Acronyms

3GPP	Third generation partnership project
5G	Fifth generation
5G NR	5G new radio
AC	Actor-critic
AD	Access delay
AD-F	Access delay counted in frames
ADMM	Alternating direction method of multipliers
AI	Artificial intelligence
AP	Access point
BFoV	Base field-of-view
BS	Base station
CNN	Convolutional neural network
CSMA	Carrier-sense multiple access
DCF	Distributed coordination function
DDPG	Deep deterministic policy gradient
DNN	Deep neural network
DQN	Deep Q network
DRL	Deep reinforcement learning
EDT	Early data transmission
EFoV	Extended field-of-view
eMBB	Enhanced mobile broadband
ET	Enhanced tile
FoV	Field-of-view
HMD	Head-mounted device
HP	High priority
IIoT	Industrial Internet of Things
IoT	Internet of Things
IoV	Internet of Vehicles
IT	Information technology
LoRa	Long range
LP	Low priority

LPWA	Low-power wide-area
LSTM	Long short-term memory
LTE	Long-term evolution
LTE-M	Long-term evolution for machine-type communications
M2M	Machine-to-machine
MAC	Medium access control
MDP	Markov decision process
mMTC	Massive machine-type communications
mmWave	Millimeter-wave
MsCS	Mini-slot based carrier sensing
MSE	Mean squared error
MTC	Machine-type communication
NB-IoT	Narrowband IoT
NOMA	Non-orthogonal multiple access
QoE	Quality of experience
QoS	Quality of service
RACH	Random access channel
RAW	Restricted access window
RMAB	Restless multi-armed bandit
RP	Regular priority
RSU	Roadside unit
SCA	Successive convex approximation
SOC	Second order cone
SMsA	Superimposed mini-slot assignment
SyncCS	Synchronization carrier sensing
TDMA	Time-division multiple access
TPSA	Task partition and scheduling algorithm
TTI	Transmission time interval
UAV	Unmanned aerial vehicle
URLLC	Ultra-reliable low-latency communications
V2I	Vehicle-to-infrastructure
V2X	Vehicle-to-everything
VR	Virtual reality
VS	Video segment
WI	Whittle index
WLAN	Wireless local area network

Chapter 1
Introduction

In this chapter, we first provide an overview of the Internet of Things from the perspectives of connected devices, use cases, deployment efforts, and technical advancement. Then, connectivity and edge computing in IoT are introduced, respectively, focusing on the requirements, available options, and challenges. The role of artificial intelligence in IoT and challenges in developing AI-based solutions are also discussed. Last, we present the scope and organization of this book.

1.1 The Era of Internet of Things

We are entering the era of the Internet of Things (IoT). Targeting to connect billions of devices, such as wearables, appliances, and industrial actuators, and a variety of systems, such as sensor networks, transportation management centers, and power grids, IoT has become a major driver worldwide for innovations in both business and technology development. The global IoT market size in 2020 is estimated to be approximately 309 billions in USD, and the forecast for 2021 and 2028 is 831 billions and 1855 billions, respectively, with an annual growth rate of 25.4% between 2021 and 2028 [1]. Meanwhile, the number of networked devices is expected to increase from around 20 billions in 2020 to almost 30 billions in 2023, with almost 15 billion machine-to-machine (M2M) connections in 2023 [2]. Moreover, it is predicted that platforms connecting devices, cloud servers, and application providers will harvest comparable revenue from emerging IoT use cases and from traditional information technology (IT) use cases by 2023 [3].

IoT is a broad concept that covers a wide range of use cases. In manufacturing industries, IoT solutions can improve asset management, optimize supply chains, and enable factory automation [4]. In agriculture, IoT platforms can facilitate plant

J. Gao et al., *Connectivity and Edge Computing in IoT: Customized Designs and AI-based Solutions*, Wireless Networks,
https://doi.org/10.1007/978-3-030-88743-8_1

status monitoring, and pest and disease control [5]. In urban management, IoT techniques can enable smart cities by integrating smart street lighting, intelligent traffic control, fire and pollution detection, etc., to promote safe, comfortable, and energy-efficient living conditions [6]. In healthcare, IoT applications can support remote in-home health monitoring for proactive and preventive diagnosis interventions [7]. In the airline business, IoT platforms can reduce fuel costs and service disruption and thereby improve customer experience [8]. Other promising IoT applications include crude oil production, wildfire detection, search and rescue, smart campus, augmented shopping, and so on [9–13].

Many countries and regions have started IoT programs or pilot projects. For example, the IoT European Large-Scale Pilots Programme has been promoting partnerships across Europe since 2016 and conducting various IoT projects with a total budget of €100 millions, including ACTIVAGE (for elderly smart living), AUTOPILOT (for automated driving), IoF2020 (for the Internet of food and farm) [14]. In the United States, New York City published its IoT strategy in March 2021 with an objective to create an IoT ecosystem for consumer, industry, and government use cases [15], while other cities, such as Las Vegas, are on course to become smart cities [16]. In China, the number of licensed IoT connections has reached 600 millions by 2018, and a major focus of future IoT development is intelligent manufacturing [17]. In addition to the above programs or pilot projects, many industry leaders have invested in and developed IoT platforms, examples of which include Amazon Web Services, Microsoft Azure IoT Platform, IBM Watson IoT Platform, and Siemens MindSphere [18].

Besides various investment from governments and industries, technology advancement in device hardware, software, communications, cloud/edge computing, artificial intelligence (AI), etc., have been propelling the development and deployment of IoT. Improvement in hardware enables the production of IoT devices with smaller sizes and lower costs [19]. Improvements in software allow IoT devices and platforms to become more secure, reliable, and energy-efficient [20, 21]. Advancement in communication technologies enables a massive scale of connections required for realizing IoT as well as new communication paradigms, such as M2M communications [22, 23]. Modern cloud and edge computing technologies provide versatile paradigms of data processing for IoT applications, allowing on-demand computing service provisioning through task offloading [24]. Lastly, advances in AI techniques render intelligent and automated connectivity and computing solutions in IoT [25].

This book focuses on the connectivity and computing aspects of IoT, with a particular focus on use case-specific designs and AI-based solutions. The rest of this chapter will discuss the basics of connectivity, edge computing, and the role of AI in IoT.

1.2 Connectivity in IoT

Connectivity is the foundation of IoT as it enables data collection from or exchange among networked IoT devices. Different network topology, connectivity requirements, and connectivity options may apply in IoT, depending on the application.

Regarding network operation, IoT applications can be implemented in a distributed, a decentralized, or a centralized manner. Examples of distributed IoT applications include distributed sensing and communication in autonomous driving [26] and plant monitoring for predictive maintenance in manufacturing [27], which require a low response time and need to process collected data locally. Examples of decentralized IoT applications include localization [28] and edge computing [29], which leverage infrastructure and resources on network edge to serve end users without relying on cloud servers. Additionally, many IoT applications adopt a client-server mode and exploit centralized cloud computing platforms, such as the enterprise platforms mentioned in Sect. 1.1. Examples of such applications are metropolitan-area intelligent transportation system planning [30] and large-scale supervisory control and data acquisition [31], which rely on extensive computing and storage resources provided by data centers.

Regarding connectivity requirements, IoT applications may require high connection density, low communication delay, long communication range, high transmission rate, or combinations of those. In a smart city scenario, 30,000 connections per square kilometer (km^2) may be needed just for connecting household water, electricity, and gas meters, which send messages with intervals between 30 min and 24 h [32]. Such connections are delay-tolerant and usually have a short range, e.g., 15 meters (m). In a factory automation setting, process state monitoring may involve 10,000 devices per km^2 [33]. Such connections span factory plants with a typical size around 300 m $\times$ 300 m $\times$ 50 m, and the delay tolerance is on the level of 50 milliseconds (ms). In internet of vehicles (IoV), a vehicle may need to simultaneously communicate with hundreds of other vehicles [34]. The connections for such communications can be transient, and the delay tolerance can be very strict, e.g., 10 ms for road safety applications.

Regarding connectivity options, various wireless communication standards and techniques are available for IoT. The three use cases of the fifth generation (5G) cellular networks, i.e., enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC), aim at providing support for various IoT applications [35]. Meanwhile, 802.11ax, or Wi-Fi 6, has enhanced support for IoT and is suitable for smart home applications [36]. In addition, a few low-power wide-area (LPWA) technologies and standards, such as Long-Term Evolution for Machine-Type Communications (LTE-M), Long Range (LoRa), Narrowband IoT (NB-IoT), support cost-effective long-range communications and are suitable for applications such as smart logistics and environment or wild-life monitoring [37]. In the future, IoT devices may also be connected via satellites.

Given the varieties of IoT applications and their connectivity requirements, finding optimal connectivity solutions is challenging, and such challenge is aggrandized when considering network heterogeneity, device mobility, network resource limitations, cost-effectiveness, and scalability. As a result, despite various potential options as mentioned above, customized designs are necessary for providing the best support to specific applications due to their unique characteristics and requirements. In Chap. 2, we will customize a connectivity solution for industrial IoT and demonstrate the potential of such customized designs for connecting IoT devices. In Chaps. 3 and 4, we will present connectivity solutions related to computing task offloading and result delivery in edge computing.

1.3 Edge Computing in IoT

Most IoT applications require not only data collection or exchange but also data analysis. As a result, they demand a computing paradigm and related resources. The data processing may happen on end user devices (such as sensors or vehicles), edge facilities (such as local network controllers), or cloud computing servers (such as Amazon Elastic Compute Cloud).

On-device processing is feasible for devices such as smartphones and vehicles, which have the hardware, software, and other resources for on-board computing [38]. Meanwhile, a significant portion of IoT devices, such as sensors and parking meters, are low-cost devices with limited processing power, storage, or battery [39]. With no or minimum on-device processing capability, such devices may resort to cloud computing and leverage resources in a cloud for data processing [40]. The cloud computing paradigm enables a variety of IoT applications and is especially suitable for applications running in a client-server mode. However, cloud computing requires devices to upload the data for processing to a cloud server, which can cause excessive traffic loads for the IoT networks when a massive number of devices rely on cloud computing. In addition, the round-trip communication, i.e., data uploading and computing result delivery, can cause a large delay that is unacceptable for applications such as autonomous driving and industrial robot arm control [41]. To reduce network traffic load and delay, edge computing has emerged as a solution, in which computing resources are deployed outside of the cloud and close to end users on network edge [42]. Such a computing paradigm is known as mobile edge computing or multi-access edge computing (MEC).

With the advent of edge computing, applications that require low-delay computing can leverage computing servers on the network edge [43]. This creates new opportunities for both IoT service providers and network operators. In smart healthcare, data collected by smartphones or wearable devices can be processed at an edge server for health monitoring applications such as gait analysis and fall risk assessment [44]. In smart cities, videos captured by cameras can be processed at edge servers for surveillance and event recognition [45, 46]. In autonomous driving, vehicles can upload data collected by cameras, radars, and other sensors to edge servers and enhance road safety via data analysis such as object recognition and tracking. In addition, many applications in various domains that leverage edge computing are emerging [47].

On the other hand, edge computing renders IoT networks more complex. New challenges arise, which often involve the synergy of computing and connectivity. For example, edge computing servers can be deployed at the access points (APs) of femtocells (e.g., home networks), small cells, and macrocells, and each deployment option has its own pros and cons [48]. In addition, the joint scheduling of transmission and computing tasks becomes critical for supporting applications with stringent delay requirements [49]. In highly dynamic networks such as vehicular networks, computing service migration or collaborative computing can be necessary for handling device mobility [50]. In Chaps. 3–5, we present edge computing solutions in representative IoT scenarios, such as IoV, and discuss various issues related to edge computing, such as task scheduling, content caching, collaborative computing, and computing result delivery.

1.4 AI in IoT

The world has witnessed a rapid advancement of AI in the past decade, with many successful real-world applications, especially in the field of natural language processing and computer vision [51]. Such success inspires the investigation on potential applications of AI in IoT, and many ideas have emerged for various use cases, such as mining, healthcare, and transportation [52, 53].

Incorporating of AI in IoT is natural. First, involving a massive number of devices, diverse applications, and spatiotemporally-variant service demands, IoT networks are complex and dynamic. AI potentially offers a viable alternative approach to managing IoT networks with the desired scalability and adaptability, while satisfying diverse and often stringent application requirements. Second, the effectiveness of AI relies on abundant data, e.g., for training neural networks, while a massive number of IoT devices can generate or provide a massive amount of data to fuel AI. Last, AI methods are suitable for data analysis in many IoT applications, such as health monitoring and fault pattern identification in smart grids [54].

AI can play a multifarious role in IoT, in terms of both the connectivity and the edge computing. Specifically, AI can be used for network traffic load prediction to facilitate IoT network planning [55]. AI can also be adopted in medium access control (MAC) to enhance IoT network throughput or fairness [56]. In addition, AI can be applied to handle computing task scheduling [57], offloading [58], and migration [59] for effective edge computing with minimum computing delay, balanced computing load distribution, or adaptivity to network dynamics.

Despite a tremendous potential of AI in empowering various IoT applications, many challenges exist in AI-based solutions for IoT. Specifically, choosing appropriate AI methods for considered IoT applications, while taking practicality into account, is essential yet challenging. Moreover, AI functionality deployment, com-