Ph.D. Thesis

A Deep Reinforcement Learning Paradigm for MAC layer resource allocation in dense Wireless Networks

Graduate School of Yeungnam University

Department of Information and Communication Engineering

Major in Information and Communication Engineering

Rashid Ali

Advisor: Sung Won Kim

February 2019

Ph.D. Thesis

A Deep Reinforcement Learning Paradigm for MAC layer resource allocation in dense Wireless Networks

Advisor: Sung Won Kim

Presented as Ph.D. Thesis

February 2019

Graduate School of Yeungnam University

Department of Information and Communication Engineering
Major in Information and Communication Engineering

Rashid Ali

Rashid Ali's Ph.D. Thesis is approved

Committee member	Prof. Seung-Yeob Nam	Seung Yeob Nam
Committee member	Prof. Sung Won Kim	
Committee member	Prof. Jin Ghoo Choi	Jin-Ghoo Choi
Committee member	Prof. Heejung Yu	
Committee member	Prof. Kwonhue Choi	

February 2019

Graduate School of Yeungnam University

Acknowledgements

In the name of almighty Allah, the most beneficent, the eternally merciful. All praises to Allah for the strength

I would like to express my sincere gratitude to my advisor **Prof. Sung Won Kim** for the patient guidance, encouragement and continuous support throughout my Ph.D. study and research. His valuable comments and reviews have played an important role to conduct this research in a proper direction as well as his guidance helped me all the time in my research and writing of this dissertation. In fact, he had been like a father taking care of his child for best learning. In addition, I would like to thank **Prof. Byung-Seo Kim** from Hongik University, Republic of Korea for his humble co-supervision. I would like to thank the rest of my dissertation committee: **Prof. Seung-Yeob Nam**, **Prof. Kwonhue Choi**, **Prof. Heejung Yu** and **Prof. Jin-Ghoo Choi** not only for their time and extreme patience, but also for their intellectual contribution to my development as a researcher. I always feel proud to have my teachers like Prof. Young-Tak Kim, Prof. Heejung Yu and Prof. Gyanendra Joshi from department of Information and Communication Engineering, Yeungnam University. I would like to special thank the people and government of Republic of Korea, for providing funding through creative human resource development team for ICT-based smart devices Brain Korea 21 Plus (BK21 Plus) and National Research Foundation (NRF).

I am very thankful to one of my best friend and colleague, Nurullah Shahin from Bangladesh for his continuous moral and technical collaboration during my Ph.D. research. My seniors and colleagues; Rojeena Bajracharya, Arslan Musaddiq, Yazdan and Nauman in the Wireless Information and Networking Lab. (*WINLab.*) and the staff of the Department of Information and Communication Engineering, thanks for your help and support. I would like to mention that my respected colleagues and friends; Dr. Yousaf Bin Zikria, Dr. Muhammad Zeeshan Jahandir, Dr. Tenvir Ali, and Muhammad Usman Ali, who have been for me a great source of motivation, not only achieving my research goals, but also my character building. I am very happy to thank Yeungnam University International Students Association (YUISA) for selecting me as a General Secretory to serve YUISA.

I think it would be incomplete if I do not acknowledge those mentors who polished me during my Masters and undergraduate studies. I am very thankful to **Prof. Zia-u-din Zia** from Gomal University, D.I. Khan Pakistan and **Dr. Stanislav Belenki** from University West, Sweden for supervising my BIT and Master in Computer Science final projects, respectively. My special gratitude is for **Dr. Maria Spante** from University West, Sweden, who has been more than a

supervisor for me during my Masters in Informatics studies. I believe that she is the only reason who inspired me for my Ph.D. studies. **Prof. Per Flensburg** from University West, Sweden is one the best mentor I had in my institutional life. His teaching philosophy is the key to my success in the field of teaching.

I would like to thank Pakistani Students community in YU for their love and care during my stay at YU. I will always remember the prayers of the Pakistani Community members; Mehmood Tufail, Amjad Ali, Muhammad Rafiq, Dr. Arif Hussain, Dr. Mehmood Arif, Dr. Imran Ashraf, Muhammad Saeed Akhtar, Waqas Khalid, Abdul Wadood Afridi, Noor Hassan, Muhammad Abdul Qyyum, Junaid Haider, Syed Naved Raza, Nadeem Jawad Haider, and Muhammad Zahoor. My special thanks to Indian community members; Dr. Arif Kashmiri, Dr. Shafique Rehman, and Tahir Lone. I am also very thankful to my brothers from Kyungpook National University, and Hayang Islamic Center, Korea, who made my educational journey as a spiritual journey as well.

My parents, who formed my current life, for their love and support, my brothers, and my sisters for supporting me throughout my life. Special thanks to my eldest brother, Syed Mumtaz Ali. I do remember the first day when I went to school with my eldest cousin Syed Mukhtyar Ali, I am very thankful to him, because he placed the first brick of my educational empire.

Most importantly, my wife to whom my warmest appreciations and thanks to for her sacrifices, support, encouragement, patience and many other things throughout this long journey. Finally, my sons, Syed Affan Ali and Syed Hamdan Ali, whose smile was my energy in this journey. They missed a lot of opportunities and time we could spend together.

Therefore, I would like to dedicate my thesis to my family, relatives, friends, and mentors, who inspired me and provided consistent encouragement during the entire process.

February 2019

Rashid Ali

Table of Contents

Dedicated to my family, relatives, friends, and mentors …

"Life is an equation of many variables,
constants and constraints,
and we are always in a wrestle of optimization."

Rashid Ali

Glossary

5G	5th generation
A-MPDU	Aggregated MAC Protocol Data Unit
AC	Access Categories
ACK	Acknowledgements
AI	Artificial Intelligence
AP	Access Point
BL	Bayesian Learning
BSS	Basic Service Set
COSB	Channel Observation-based Scaled Backoff
CFP	Contention-Free Period
CRN	Cognitive Radio Network
CS	Carrier Sensing
CSMA/CA	Carrier Sense Multiple Access with collision avoidance
CW	Contention Window
D2D	Device-to-Device
DCF	Distributed Coordination Function
DSSS	Discrete Sequence Spread Spectrum
DL	Deep Learning
DRL	Deep Reinforcement Learning
DIFS	Distributed Inter-frame Space
EDCA	Enhanced Distributed Channel Access
FDMA	Frequency Division Multiple Accesses
HCF	Hybrid Coordination Function
HEW	High Efficiency WLAN
HMM	Hidden Markov Model
H-SVM	Hierarchical SVM
ICA	Independent Component Analysis
IEEE	Institute of Electrical and Electronics Engineers
IEEE-SA	IEEE Standards Association
IoT	Internet-of-Things
***i*QRA**	*intelligent* QL-based Resource Allocation
KNN	*k*-Nearest Neighbor
MAC	Medium Access Control
MAC-RA	MAC layer Resource Allocation
MDP	Markov Decision Process
MI	Machine Intelligence

MIP	Mixed-Integer Programming
MIMO	Multiple Input Multiple Output
MU-MIMO	Multi-User MIMO
NAV	Network Allocation Vector
OFDM	Orthogonal Frequency Division Multiplexing
PC	Principal Components
PCA	Principle Component Analysis
PCF	Point Coordination Function
PFP	Polling-based Contention Process
PHY	Physical Layer
POMDP	Partially Observed MDP
QAM	Quadrature Amplitude Modulation
QL	Q Learning
QoE	Quality of Experience
QoS	Quality of Service
RA	Regression Analysis
RTS/CTS	Request-To-Send/Clear-To-Send
SDMA	Space Division Multiple Access
STA	Station
SVM	Support Vector Machine
TDMA	Time Division Multiple Access
TG	Task Group
TXOP	Transmission Opportunity
VHT	Very High Throughput
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network

List of Figures

List of Tables

Ph.D. Thesis

A Deep Reinforcement Learning Paradigm for MAC layer resource allocation in dense Wireless Networks

Rashid Ali

(Supervised by: Professor Sung Won Kim)

Graduate School of Yeungnam University

Department of Information and Communication Engineering

Major in Information and Communication Engineering

ABSTRACT

Wireless local area networks (WLANs) are widely deployed for internet-centric data applications. It is predicted that by 2020, about two-thirds of the world's internet traffic will be video, and more than half of the traffic will be offloaded to WLANs. Consequently, WLANs need major improvements in both throughput and efficiency. New technologies continue to be introduced for WLAN applications for this purpose. The IEEE 802.11ac standard is the currently implemented amendment by the IEEE 802.11 standard working group that promises data rates at gigabits per second. The main features of the IEEE 802.11ac standard are adopting increased bandwidth and higher order modulation than the previous standards, and multiple-input multiple-output (MIMO) and multi-user MIMO transmission modes. These features are designed to improve the user experience. In addition to technologies that enhance the efficiency of the WLAN, the IEEE 802.11ax

High Efficiency WLAN (HEW) standard is also investigating and evaluating advanced wireless technologies to utilize the existing spectrum more efficiently.

The next-generation dense WLAN, HEW is expected to confront ultra-dense user environments and radically new applications. HEW is likely to provide four times higher network efficiency even in highly dense network deployments. However, the current WLAN itself faces huge challenge of efficient channel access due to its temporal-based MAC layer resource allocation (MAC-RA). WLAN uses a carrier sense multiple access with collision avoidance (CSMA/CA) procedure to access the channel resources, which is based on a binary exponential backoff (BEB) mechanism. In BEB, a random backoff value is generated from a contention window (CW) to obtain channel access. The CW size is doubled after every unsuccessful transmission and reset to its minimum value on successfully transmissions. However, this blindness when increasing and resetting the CW induces performance degradation. For a dense network, resetting the CW to its minimum size may result in more collisions and poor network performance. Likewise, for a small network environment, a blind increase in CW size may cause an unnecessarily long delay while accessing the channel. To satisfy the diverse requirements of dense WLANs, it is anticipated that prospective HEW will autonomously access the best channel resources with the assistance of sophisticated wireless channel condition inference in order to control channel collisions. Such intelligence is possible with the introduction of deep learning (DL) techniques in future WLANs.

The potential applications of DL to the MAC layer of IEEE 802.11 standards have already been progressively acknowledged due to their novel features for future communications. Their new features challenge conventional communications theories with more sophisticated artificial intelligence-based theories. DL has been extensively proposed for the MAC layer of WLANs in various research areas, such as deployment of cognitive radio and communications networks. Deep reinforcement learning (DRL) is one of the DL technique that is motivated by the behaviorist sensibility and control philosophy, where a learner can achieve an objective by interacting with the environment. In this dissertation, a DRL-based intelligent paradigm is developed for MAC-RA in dense WLANs.

One of the DRL models, Q-learning (QL), is used to propose an intelligent QL-based resource allocation (*i*QRA) mechanism for MAC-RA in dense WLANs. *i*QRA exploits channel observation-based collision probability for network inference to dynamically and autonomously control the backoff parameters (such as backoff stages and CW values). The simulations performed in network simulator 3 (ns3) indicate that the proposed DRL-based *i*QRA paradigm learns diverse WLAN environments and optimizes its performance, compared to conventional non-intelligent MAC protocol, BEB. The performance of the proposed *i*QRA mechanism is evaluated in diverse WLAN network environments with throughput, channel access delay, and fairness as performance metrics.

1. Introduction

1.1 Motivation

Future dense wireless local area networks (WLANs) are attracting significant devotion from researchers and industrial communities. IEEE working groups are expected to launch an amendment to the IEEE 802.11 (WLAN) standard by the end of 2019 [1]. The upcoming amendment, covering the IEEE 802.11ax high-efficiency WLAN (HEW), will deal with ultra-dense and diverse user environments, such as sports stadiums, train stations, and shopping malls. One inspiring service is the promise of astonishingly high throughput to support extensively advanced technologies for 5th generation (5G) communications. HEW is anticipated to infer the various and interesting features of both the learners' environment of a HEW device as well as device behavior in order to spontaneously control the optimal media access control (MAC) layer resource allocation (MAC-RA) [2] system parameters.

In real WLANs, the devices proficiently and dynamically manage WLAN resources, such as the MAC layer carrier sense multiple access with collision avoidance (CSMA/CA) mechanism to improve users' quality of experience (QoE) [3]. Overall device performance depends on exploitation of the instability of network heterogeneity and traffic diversity. WLAN resources are fundamentally limited due to shared channel access and wireless infrastructures, whereas WLAN services have become increasingly sophisticated and diverse, each with a wide range of QoE requirements. Thus, for the success of the prospective HEW, it is vital to investigate efficient and robust MAC-RA protocols [2].

The main motivation for the work in this thesis is to highlight the issues and challenges in MAC-RA for upcoming HEW wireless networks. The aim of the thesis is to increase the efficiency and robustness of the MAC layer in next generation wireless networks (that is HEW). In order to infer the diverse and

interesting features of users' environments as well as users' behaviors, future HEW must spontaneously control optimal MAC-RA system parameters. Recently, the field of deep learning (DL) has been flourishing in order to enable machine intelligence (MI) capabilities in wireless communications technologies. It is believed by researchers that WLANs can optimize performance by introducing DL into MAC layer resource allocation. Deep reinforcement learning (DRL) is one DL technique that is motivated by the behaviorist sensibility and control philosophy, where a learner can achieve an objective by interacting with the environment [4]. DRL uses specific learning models, such as the Markov decision process (MDP), the partially observed MDP (POMDP), and Q-learning (QL) [5]. DRL utilizes these techniques in applications like learning an unknown wireless network environment and resource allocation in femto/small cells in heterogeneous networks (HetNets) [5].

In this thesis one of the models of DRL, Q learning is employed to acquire awareness about the state evaluation of the MAC-RA in wireless networks. QL-based intelligent MAC-RA optimizes the performance of wireless networks especially in dense network environments.

1.2 Scope of the thesis

As discussed earlier, the main objective of this thesis is to develop theory and method for learning how MAC-RA channel access of the WLANs evolves and for identifying optimal parameters selection in an intelligent manner. To these ends, two specific goals are identified, which will be described next.

The first goal of this thesis is to study and quantify the issues and challenges for MAC-RA in WLANs, more specifically in dense wireless networks (published in Publication I). It is found that the increase in density of network devices is directly proportional to the collision in the network, and results in network performance degradation. The traditional CSMA/CA channel access mechanism is evaluated for this purpose. The binary exponential backoff (BEB) algorithm of

CSMA/CA blindly handles the collision avoidance in the traditional WLANs (for detailed discussion please refer to the Section 2.4 “Problem Statement”). A practical channel observation-based scaled backoff (COSB) algorithm, that overcomes the blindness issue of BEB, is proposed in Publication II and Publication III of Appendix A.

The second goal of the thesis is to develop method for network device’s experience inference in order to spontaneously control the optimal MAC-RA parameters of proposed COSB. In this thesis, the potentials of deep reinforcement learning paradigm are studied for performance optimization of channel observation-based MAC protocols (that is COSB) in dense wireless network. An *intelligent* Q learning-based resource allocation (*i*QRA) mechanism is proposed for this purpose, where Q learning is one of the prevailing models of DRL. The *i*QRA mechanism is published in Publication IV.

1.3 Contributions of the thesis

The main contributions of this thesis are presented in the original Publications I-IV of the Appendix A. In details, the contributions of this thesis are:

- A practical channel observation-based scaled backoff (COSB) is proposed in Publication II-III. It is illustrated that a practical channel collision probability can proficiently be measured by observing the channel status that is busy or idle. The measured channel collision probability is utilized to scale the backoff contention window (CW) in CSMA/CA. The proposed COSB mechanism enhances the performance of CSMA/CA. An analytical modeling for the proposed COSB is also formulated in Publication III to affirm the validity of the proposed mechanism.
- A deep reinforcement-learning paradigm is proposed for performance optimization of COSB algorithm. The proposed paradigm utilizes Q learning, one of the prevailing deep reinforcement learning model for

network inference. QL algorithms are useful to optimize the performance of a device through experience gained from the interaction with the unknown environment. QL is inspired by behaviorist psychology, and is used to discover an optimum strategy for taking action from any finite Markov decision process (MDP), mainly when the environment is unknown. A significant feature of QL is off-policy, which explicitly considers the problem of a target-oriented learner, which continuously interacts with an uncertain (model-free) environment. All QL learners have explicit targets, such that they can sense features of their environment and can choose actions to affect their environment. A target-oriented learner can be an element of a larger behavioral system, such as a HEW device in a WLAN environment seeking to maximize its performance. Because QL finds solutions through the experience of interacting with an unknown environment, it is used to optimize the CW size adjustment in the COSB mechanism. An *intelligent* QL-based resource allocation (*i*QRA) mechanism is proposed to optimize the performance of the COSB mechanism for dense HEW networks. By using network inference, *i*QRA dynamically and autonomously controls backoff parameters (i.e. the backoff stages and CW sizes) selection in COSB.

1.4 Summary of publications and author's contributions

Publication I (Abstract): It is predicted that by 2020, about two-thirds of the world's internet traffic will be video and more than half of the traffic will be offloaded to Wi-Fi networks. Consequently, WLANs need major improvements in both throughput and efficiency. New technologies continue to be introduced for WLAN applications for this purpose. The IEEE 802.11ac standard is the currently implemented amendment by the IEEE 802.11 standard working group that promises data rates at gigabits per second. The main features of the IEEE 802.11ac

standard are adopting increased bandwidth and higher order modulation than the previous standard, and multiple-input multiple-output (MIMO) and multi-user MIMO transmission modes. These features are designed to improve the user experience. In addition to technologies that enhance the efficiency of the WLAN, the IEEE 802.11ax standard is also investigating and evaluating advanced wireless technologies to utilize the existing spectrum more efficiently. These modern communications technologies are steadily advancing physical layer data rates in WLANs, although data throughput efficiency of the WLAN may degrade rapidly as the physical layer data rate increases. The fundamental reason for the degradation is that the current MAC protocol allocates the entire channel to one user as a single source due to equally distributed time domain contention resolution. The challenges and difficulties have already been identified for designing efficient MACRA schemes for the upcoming IEEE 802.11ax HEW networks. However, there is no profound investigation outcome for this kind of efficient resource allocation. Therefore, in this publication, an extensive survey of the expected features and challenges for IEEE 802.11ax in the design of fair and efficient MAC-RA is conducted. The associated previous research work is summarized as to future directions. Moreover, the need for each directed scheme is highlighted.

Publication II-III (Abstract): The IEEE 802.11ax HEW is promising as a foundation for evolving the 5G radio access network on unlicensed bands (5G-U). 5G-U is a continued effort toward rich ubiquitous communication infrastructures, promising faster and reliable services for the end user. HEW is likely to provide four times higher network efficiency even in highly dense network deployments. However, the current WLAN itself faces huge challenge of efficient channel access due to its contention-based nature. WLAN uses the CSMA/CA procedure in MAC protocols, which is based the BEB mechanism. Blind increase and decrease of the contention window in BEB limits the performance of WLAN to a limited number of contenders, thus affecting end-user quality of experience. In these publications, the future use cases of HEW for 5G-U networks are identified. Further, a self-

scrutinized channel observation-based scaled backoff (COSB) mechanism is proposed to handle the high-density contention challenges. A recursive discrete-time Markov chain model (R-DTMC) is formulated to analyze the performance efficiency of the proposed COSB mechanism.

Publication IV (Abstract): Next-generation wireless networks for 5G and Internet of Things (IoT) systems are expected to confront ultra-dense user environments and radically new applications. To satisfy the diverse requirements of these future technologies, wireless networks will face the challenge of assisting MAC-RA in intelligent adaptive learning and decision-making. Deep learning (DL) offers services as an auspicious artificial intelligence tool for wireless-enabled IoT devices. It is expected that future wireless systems will independently access the most commendable channel resources with the assistance of sophisticated wireless channel condition inference in order to control channel collisions. Intelligent paradigms can simultaneously adjust channel contention parameters with the aid of channel collision probability inference while relying on throughput-efficiency learning.

Therefore, in these publication, DRL is proposed as an intelligent paradigm for MAC layer resource allocation in dense WLANs. One of the DRL models, Q-learning (QL), is used to optimize the performance of channel observation–based MAC protocols (such as COSB) in dense wireless networks. An *intelligent* QL-based resource allocation (*i*QRA) mechanism is proposed for MAC layer channel access in dense WLANs. Simulation results indicate that the proposed intelligent paradigm *i*QRA learns diverse WLAN environments and optimizes performance, compared to conventional non-intelligent MAC protocols. The performance of the proposed *i*QRA mechanism is evaluated in diverse WLAN network environments with throughput, channel access delay, and fairness as performance metrics.

The author of this thesis is responsible for the derivations, algorithm development and simulations in all the publications of this thesis. The author of this thesis has also been in charge of writing the first drafts of the papers. The co-

authors have provided insightful help in planning and revising the papers. The co-authors also gave crucial suggestions for the design of the algorithms proposed in Publications I-III and for the design of the deep learning paradigm proposed in Publication IV.

1.5 Organization of the thesis

This thesis consists of six chapters that introduce the six original publications attached to the Appendix-A (Publications I-IV) at the end of this thesis. Chapter 2 gives an overview of IEEE 802.11 technologies and challenges related to MAC layer resource allocation in IEEE 802.11 networks. The main contributions of this thesis are presented in Chapters 3–5. Chapter 3 identifies the potentials of deep learning in wireless networks. Different types of deep learning are discussed with their applications in wireless networks. Deep reinforcement learning is proposed as a paradigm for MAC layer resource allocations in wireless networks. For this purpose, Q learning which is one of the models of deep reinforcement learning is proposed as a MAC-RA paradigm. The type and models of deep learning used in this thesis are defined together with a compact survey of the deep learning paradigm in wireless networks proposed in the literature. In Chapter 4, the *intelligent* Q learning-based resource allocation (*i*QRA) is presented. This chapter also includes the proposed channel observation-based scaled backoff (COSB) mechanism. Later in this chapter, COSB is optimized with the integration of *i*QRA mechanism. In Chapter 5, the performances of the proposed deep reinforcement learning-based mechanism, that is *i*QRA is evaluated using multiple performance metrics. Concluding remarks and discussion about the future research directions are given in Chapter 6.

2. Preliminaries

WLANs are experiencing extensive growth in internet-centric data applications. Advanced technology markets are utilizing WLANs, and deployments are rapidly flourishing in public and private areas, like shopping malls, cafes, hotels and restaurants, bus/train stations, airports, etc. In addition, there is the rapid increase of WLAN-enabled electronic devices, because consumers demand that their entertainment devices be internet-enabled. In order to cover new device categories and new applications, exciting new technologies are emerging for WLANs in order to address the need for increased network capacity and coverage, efficient energy consumption, and ease of use. Consequently, WLANs need major improvements in both throughput and efficiency. New technologies for WLAN applications are continuously introduced. The IEEE standard for WLANs was initiated in 1988 as IEEE 802.4L [6], and in 1990, the designation changed to IEEE 802.11 to form a WLAN standard. This standard describes the PHY layer [6] and MAC sub-layer specifications for portable, stationary, and mobile devices within a local area for wireless connectivity. The IEEE 802.11ac [7] standard is the currently implemented amendment from the 802.11 standard working group (WG) promising data rates at gigabits per second. These modern communications standards and technologies are steadily advancing PHY layer data rates in WLANs. This capacity growth is achieved primarily through increased channel bandwidths and advanced PHY layer techniques, like multiple-input multiple-output (MIMO) and multi-user MIMO (MU-MIMO). These modern communications technologies are advancing PHY layer data rates in WLANs, although data throughput efficiency in WLANs may degrade rapidly as the PHY layer data rate increases. The fundamental reason for this degradation is that the current random access-based MAC protocol allocates the entire channel to one user as a single source due to equally distributed time domain contention resolution. Even if senders have a small amount (or less critical) data to send, they still need to contend for the entire channel and get an equally distributed time opportunity for transmission. As a result, the higher the PHY layer data rate,

the lower the throughput efficiency achieved. The strategies like channel bonding, frame aggregation and block acknowledgment, reverse direction forwarding, etc. enhance the high throughput capabilities in 802.11 MAC protocol [7-8]. IEEE 802.11 standard-based WLANs often struggle to service diverse workloads and data types. Since the applications are categorized into different priorities by the access layer protocol, the method how to provide enhanced and efficient resource allocation has become an interesting and challenging topic.

Although the researchers have spent plenty of time on 802.11 MAC protocol throughput enhancements using above mentioned techniques, efficient medium allocation in the MAC layer is still one of the important target areas for future WLAN researchers. Therefore, in this chapter, first an overview of standards and amendments of IEEE 802.11 MAC protocols is discussed. Later, after a brief discussion of MAC layer resource allocation (MAC-RA) in IEEE 802.11, a problem statement is formulated.

2.1 Overview of IEEE 802.11 MAC protocols

Since the inception of the WLAN in 1990, with its technical specifications rooted in the current IEEE 802.11 standard, most of the progress made has been for higher data throughput. It took many years for IEEE 802.11 WGs to approve the first draft and to later evolve it into its many standards and amendments, particularly for higher speed PHY layer transmission [9]. The terms standard and amendment can be used interchangeably; however, more precisely standards are documents with mandatory requirements and amendments are the documents that add to, remove from in a portion of existing standards. These were designated as IEEE 802.11b, 11a/g, 11n, and 11ac. For enhancement of quality of service (QoS), there is 802.11e; for security, 802.11i; for wireless access in vehicular environments, 802.11p; and to describe mesh networking, 802.11s. In the following sub-sections, we focus on the background information of some of the important standards/amendments that are closely related to the topic of the paper. Figure 2.1 summarizes a fictional journey

of IEEE 802.11 standards until currently implemented 802.11ac amendment. Table 2.1 highlights and summarizes some of the important standards and amendments.

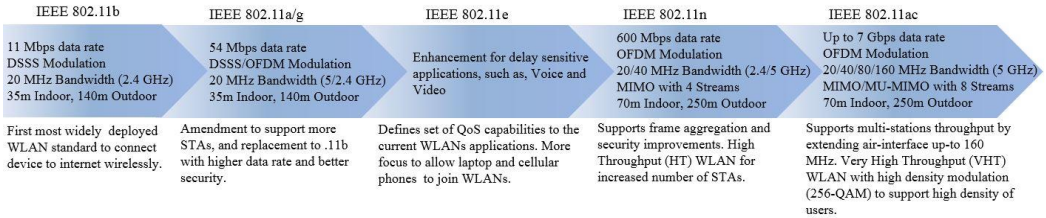

Figure 2-1 Fictional journey of 802.11 standards until 11ac amendment.

Table 2.1 Popular 802.11 amendments.

Amendments	Purpose
802.11 (1997)	Original release, data rate up to 2 Mbps in 2.4 GHz ISM band based on DSSS PHY
802.11b (1999)	Extends the DSSS PHY layer, enabling up to 11 Mbps data rates in 2.4 GHz ISM band; high-speed physical layer extension
802.11a (1999)	Adds an OFDM-based PHY layer at 5 GHz, enabling 54 Mbps to be achieved in ISM band
802.11c (2001)	For proper bridge operations, mainly for AP deployment
802.11g (2003)	Defines higher data rate PHY layer extension up to 54 Mbps at 2.4 GHz (backward compatable to 802.11b)
802.11h (2004)	Covers Dynamic Frequency Selection (DFS) and Transmit Power Control (TPC); allows 802.11a for European compatibility
802.11i (2004)	Includes improved security; replaces previous Wired Equivalent Privacy (WEP) security specification
802.11e (2005)	Enhaces the 802.11 MAC protocol to support QoS, including packet bursting
802.11r (2008)	Faster and secure handoff from one base station to another; managed in a seamless manner
802.11n (2009)	Improves standard by MIMO in both 2.4 GHz and 5 GHz bands to maximize data rate from 54 Mbps to 600 Mbps; improves security
802.11p (2010)	Wireless Access for the Vehicular Environment (WAVE), it defines architecture and series of services
802.11s (2011)	To create the mesh topologies of wireless networks, Extended Service Set (ESS)
802.11ad (2012)	Provides high theoretical throughput (7 Gbps) in 60 GHz band
802.11ac (2013)	Extends 802.11n in 5 GHz band to offer higher throughput, higher density modulation, and additional MIMO streams to give a theoretical throughput of 7 Gbps in bands < 6 GHz
802.11ah (2015-2016)	Extends range, making it useful for rural communications and offloading cell phone tower traffic; expected to be finalized and implemented in 2016
802.11ax (2019)	Expected to be approved in 2018-19, a high-efficiency WLAN (HEW) that will modify both 802.11 PHY layer and 802.11 MAC protocol, especially for densely deployed environments

2.1.1 IEEE 802.11b Amendment

The IEEE 802.11b amendment to the original standard (IEEE 802.11-97) was endorsed in 1990. It is the most frequently used WLAN extension of the discrete

sequence spread spectrum (DSSS) at 2.4 GHz to achieve data rates of 5.5 and 11 Mbps. The rate shift mechanism of 802.11b makes it possible for high data rate networks to slow the rate down to 1 or 2 Mbps. The techniques like channel bonding and the researches to increase the speed up to 22/33/44 Mbps introduced burst transmission. The IEEE 802.11b uses the same CSMA/CA resource allocation schemes defined in the original standard [6].

2.1.2 IEEE 802.11a/g Amendment

This amendment was also ratified in 1999, to use the same core protocol as in original standards. It operates in 5 GHz frequency band, and uses an orthogonal frequency division multiplexing (OFDM) with a maximum data rate of 54 Mbps. Similar to the 802.11b, the shifting of rate can help to reduce the data rate to 48/36/34/18/12/9 and 6 Mbps under lossy environments. The fundamental distributed coordination function (DCF) medium access scheme, which relies on the CSMA/CA, was carried on by the 802.11a amendment. In June 2003, another modulation standard was introduced with letter “g”. This amendment used the 2.4 GHz band (like in b), and can operate in 5 GHz band (like in a) with the maximum data rate of 54 Mbps. Therefore, it is fully backward compatible with both, 11b and 11a. The presence of 802.11b STAs in the WLAN significantly reduces the speed of an 802.11g STA due to the use of DSSS and OFDM. This backward compatibility in 802.11g can be considered as disadvantage. Furthermore, the drawback of 802.11g involves the complexities of implementation since the later amendments involves a less complex implementation.

2.1.3 IEEE 802.11e Amendment

In 2005, IEEE developed a new extension of 802.11 to support QoS provisioning in WLAN. The extension was to overcome the problem tempted by the MAC layer techniques like point coordination function (PCF) and DCF in the earlier extensions. With PCF, a centralized device allocates the resources to the STA in the network [7]. Since such centralized devices are not available in certain cases, a distributed

scheduling scheme is expected to be implemented. Due to the best effort service nature of DCF, it is unable to provide efficient performance for voice and video applications in WLANs. A hybrid coordination function (HCF) is introduced, which combines the features of PCF and DCF for resource allocation in 802.11. IEEE 802.11e is one of the WLAN standards providing QoS for voice and video applications using HCF, along with enhanced distributed channel access (EDCA). In addition to these, two more MAC enhancements were also introduced in 802.11e to improve the MAC layer throughput; the Block Acknowledgement which enabled sending of a single acknowledgment (ACK) for a block of frames; and Direct Link Protocol, which enabled direct link between two STAs in a single WLAN network.

2.1.4 IEEE 802.11n Amendment

In 2004, IEEE announced a new TG to develop a new amendment to the 802.11 standard. The purpose was to increase the real data throughput to at least 100 Mbps, which may require even high raw data rate at the PHY layer to come up with 4–5 times faster than the 802.11a or 802.11g, and 20 times faster than 802.11b. The IEEE 802.11 TG announced high throughput 802.11n in 2009, improving standard by MIMO in both 2.4 and 5 GHz bands to maximize the data rate from 54 to 600 Mbps. The IEEE 802.11n MAC enhancements are aimed at overcoming the inefficiencies of the legacy 802.11 MAC while preserving the backward compatibility. The enhancements such as adaptive coordination function, an extension to HCF and EDCA, frame aggregation, and closed loop operation were made to make MAC layer highly efficient. The techniques like block ACK and reverse direction forwarding were introduced to enhance the efficiency of transmission opportunity (TXOP). The security was also improved as compared to the legacy amendments.

2.1.5 IEEE 802.11ac Amendment

The latest amendment to the journey of 802.11 is the 802.11ac, approved in 2013. It is designed to work exclusively in the 5 GHz band. This amendment was driven

by the need for high speed due to rapid increase in the use of internet-centric devices. IEEE 802.11ac aims to provide an aggregated throughput of up to 1 Gbps, namely very high throughput (VHT) WLANs. This significant improvement is achieved by introducing novel PHY and MAC layer features, such as the use of 80 and 160 MHz channel bandwidths. A denser modulation scheme 256-quadrature amplitude modulation (QAM), an aggregated MAC protocol data unit (A-MPDU), and most importantly, the support for MU-MIMO to support simultaneous transmission of up to four STAs in the maximum of eight streams are introduced. Most of the enhancements were made to PHY layer in 802.11ac amendment and MAC layer is mostly modified to adapt to these PHY layer changes. The introduction of SDMA resource allocation for MU-MIMO has increased 802.11 MAC layer's efficiency to multiple folds. Many 802.11 devices are battery powered, thus power saving enhancements are worthwhile. IEEE 802.11ac introduced VHT TXOP power-save feature, in which an STA can switch off its radios after knowing that AP has assigned TXOP to another STA.

2.1.6 IEEE 802.11ax High Efficiency WLAN (HEW)

The IEEE Standards Association (IEEE-SA) approved standardization activity of the IEEE 802.11ax TG, concerned with densely deployed WLANs, in May 2014 [1]. Calling it HEW, the scope of the IEEE 802.11ax amendment is to define modifications for both the 802.11 PHY and 802.11 MAC layers that enable at least four-fold improvement in the average throughput per station in densely deployed networks. It is also assumed that it shall provide backward compatibility with existing IEEE 802.11 devices operating in the same band. Unlike previous amendments, this one focuses on improving metrics that reflect the user experience. The desired enhancements will be made to support dense environments, such as wireless corporate offices, outdoor hotspots, residential apartments, and stadiums. The actual deployment of the standard is anticipated for late 2019 [10]. Figure 2.2 illustrates the possible timeline and progress towards the IEEE 802.11ax standard.

The main emphasis in IEEE 802.11ax is to improve real-world performance as experienced by end users by enhancing the per-STA throughput. Possible approaches deal with three major problems in dense WLAN environments: congestion, interference, and frame conflicts. Figure 2.3 shows few of the technologies discussed in the IEEE 802.11ax TG [11].

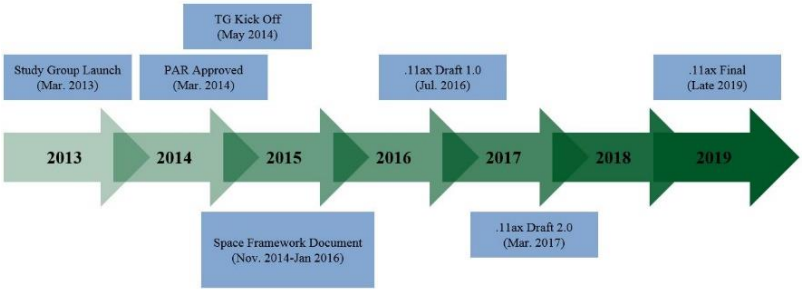

Figure 2-2 Possible IEEE 802.11ax timeline and progress.

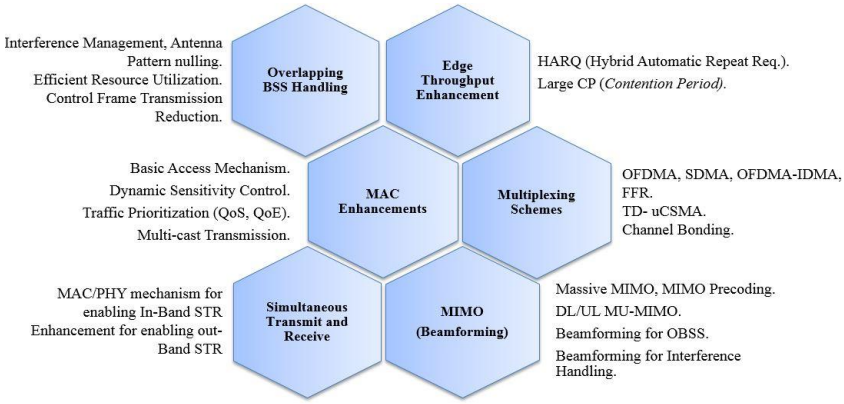

Figure 2-3 Technologies discussed in the IEEE 802.11ax study.

2.2 Overview of MAC layer resource allocation in IEEE 802.11

The MAC layer resources can mainly be allocated in three dimensions: frequency-based allocation, spatial-based allocation, and temporal-based allocation. These are known as frequency division multiple accesses (FDMA), space division multiple access (SDMA), and time division multiple access (TDMA) mechanisms.

2.2.1 Frequency-based MAC-RA

One of the emerging technologies to improve the MAC resource allocation efficiency is OFDMA. The OFDMA technology is used to divide whole channel into several sub-channels, and sub-channels are further divided into several sub-carriers. It enables multi-user channel access and multi-user data transmission since different STAs could use different sub-channels simultaneously. The associated researchers [2] have proved that the introduction of OFDMA into 802.11 MAC protocol makes remarkable improvements for efficiency. Even if all STAs are capable of transmitting on the entire available frequency, the current 802.11 MAC protocol allows the AP to transmit to, or receive from, only a single STA in a time slot. The latest amendment IEEE 802.11ac allows the AP to transmit to multiple STAs simultaneously in the downlink using the OFDMA. Although the OFDMA is an efficient candidate for the future WLANs [12], it still has problems due to the channel contention and backoff that have to be performed for each transmission thus causing excessive overhead. When an AP has packets to transmit to more than one STA, it has to compete for the medium at least total number of receiving STAs before it can send out packets. Another challenge is to maintain QoS, especially in densely deployed scenarios, where relatively large number of STAs in a single basic service set (BSS) exists having packets with high QoS requirements. The throughput of WLANs can be improved by taking advantage of multi-user (MU) diversity in channel frequency-based domain. The challenge is to select the specific STAs that will take part in the uplink MU transmission. Therefore, the design of an efficient mechanism to create groups of STAs with low channel correlation and similar channel quality is still an open issue for the next-generation WLANs.

2.2.2 Spatial-based MAC-RA

Multiple STAs can transmit different information in different physical areas, known as the spatial separation. The spatial-based resource allocation utilizes the physical space separation of the users in order to optimize the use of spectrum. In

the space domain, the transmit power of each STA is controlled by SDMA [13]. SDMA serves different STAs by using spot beam antenna technology [14]. These areas may be served by the same frequency or different frequencies. There is comprehensive work done by the researchers [2] to show the importance of spatial-based resource allocation. However, it is required that the WLAN coverage cells are sufficiently separated due to the co-channel interference. It limits the number of coverage cells, hence limits the frequency re-use factor. More advanced method combined with other resource allocation schemes (frequency domain and time domain) can further increases the capacity of the network.

2.2.3 Temporal-based MAC-RA

In wireless communication, continuous transmission is not required because STAs do not use the allocated bandwidth all the time. In such cases, temporal-based access technique is favorable to 802.11 WLANs. It is observed that temporal-based resource allocation is the dominant and most important resource allocation in 802.11 WLAN. The reason is the randomness and distributed nature of the STAs in the WLAN. Although the frequency-based MAC-RA and spatial-based MAC-RA allow multi-user uplink and downlink transmissions in WLANs, which is one of the promising techniques of the future WLANs, these schemes are still bound to be followed by time domain to transmit at the same time.

2.3 MAC-RA coordination functions

Figure 2.4 describes the MAC layer medium access coordination functions used for temporal-based resource allocation in WLANs. These functions are mainly divided into two categories: contention-based (random channel access) and contention-free (fixed assignment channel access). Both categories differ in the topological structure of the network, as well as in coordination function. These categories are further divided into DCF, EDCA, PCF, and HCF.

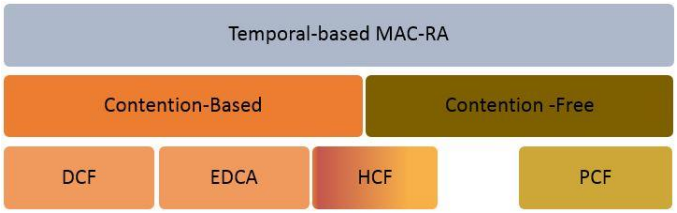

Figure 2-4 Temporal-based MAC-RA coordination functions.

2.3.1 Contention-based coordination functions

The initial IEEE 802.11 standard defines a contention-based distributed medium access algorithm known as DCF. The DCF uses CSMA/CA to contend for channel access. DCF can either operate under the basic access scheme [Figure 2.5(a)] or the optional request-to-send/clear-to-send (RTS/CTS) [Figure 2.5(b)] scheme. RTS/CTS access scheme is introduced to resolve the hidden node problem in WLANs [2]. Some STA's transmissions are not detected during carrier sensing (CS) by other STAs, but those transmissions interfere with transmission of other STAs. These STAs are hidden from each other and can cause collisions due to unawareness of the medium access conditions. In RTS/CTS, STA transmits RTS packet and wait to receive CTS packet before actual data transmission, as shown in Figure 2.5(b). Binary exponential backoff (BEB) and a deferral mechanism are used to differentiate the transmission start time of each station. BEB is used by STAs to contend with other STAs to access a medium and to transmit data. It is defined as the discrete backoff time slots for which the STA has to defer before accessing the medium. The BEB mechanism is initiated after the channel has been found idle for a predefined distributed inter-frame space (DIFS) period. Other STAs overhear the transmission from neighboring STAs by CS and set up their network allocation vector (NAV) to avoid collisions. The applications over the network are categorized into different priorities by the access layer protocol, but DCF does not explicitly support QoS for specific traffic sessions with QoS requirements. The STAs with identical back-off and deferral values can grab the whole bandwidth from QoS-sensitive sessions.

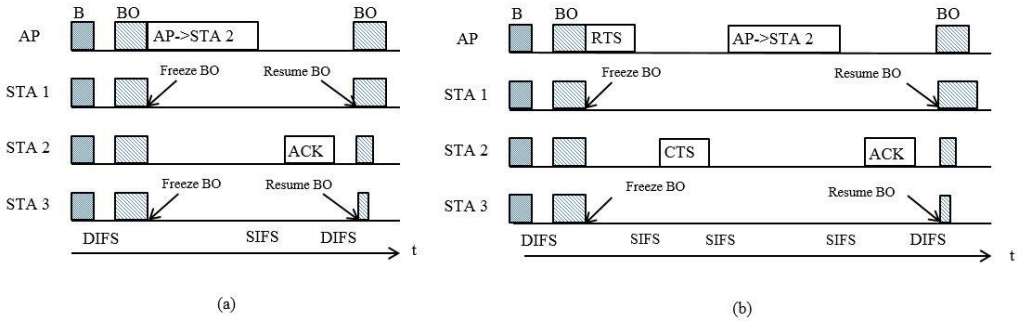

Figure 2-5 IEEE 802.11 DCF transmission procedure: (a) basic access mechanism and (b) RTS/CTS mechanism.

Similarly, in case of high traffic load and density, collisions will increase dramatically for the contending nodes. EDCA is used to handle such multi type of service communications in wireless networks. The main difference between DCF and EDCA is that DCF possesses only one queue for all types of data, whereas EDCA divides the coming packets into four types of logical queues, known as access categories (ACs). These ACs are defined according to the prioritized applications. The higher priority data is assigned a shorter backoff and deferral duration to obtain more chances to access the medium than other types of session. If any collision happens among more than one AC, the higher priority AC secures the opportunity to access the channel, while the others restart their backoff processes [15]. This differentiated access service for different types of application still suffers from degraded performance under densely deployed traffic situations where many contending STAs are present. That is because EDCA still follows contention-based medium access, and collisions are still possible. Moreover, high priority applications hardly provide low priority traffic any opportunity to access the resources. EDCA has this kind of unfairness. The effectiveness of EDCA suffers from more issues in the presence of hidden terminals and interference [16].

2.3.2 Contention-free coordination functions

PCF is different from the distributed medium access of DCF and EDCA, where all STAs communicate with each other via a centralized STA called a point

coordinator, which usually resides in the AP. The PC controls the access period of the medium by splitting the resource airtime into super-frames of contention-free periods (CFPs). The controlled access period still follows the polling-based contention process (PFP). In the CFP, each STA initially sets its NAV to the maximum duration (CFPMaxDuration) at the beginning of each CFP. This NAV duration is reset if the STA receives a CF-End or a CF-End plus ACK frame from the AP, indicating the end of the CFP. Although the PCF is a contention-free medium access in the CFP, there are still several issues with providing efficiency like QoS for applications, because it does not support AC differentiation for priority applications. There is no prediction for the occupancy of the resource by the polled STA. Therefore, the fairness issue persists in the network for each STA, because the transmission time is not bounded. With densely deployed networks, this problem can cause more severe unfairness [17].

2.4 Problem Statement

While future physical-link technologies promise to deliver sufficient bandwidth to serve user demands, existing CSMA-based channel access schemes under IEEE 802.11 are inefficient for large numbers of STAs with extensively changing demands. However, the efficiency of the current medium access protocols will soon encounter challenges when networks are deployed even more densely, like a network having to support thousands of users (STAs), or access points (APs) deployed in very close proximity to each other. Such that a stadium, a train or an apartment building where the density of WLAN users is very high. Most of the challenges come with the efforts to implement MAC-RA in distributed types of wireless network, specifically when there is no centralized station controlling the dedicated resource allocation and disseminating the reservation control information. One of the issues with proposing MAC-RA schemes is how to efficiently allocate available resources to STAs. A proper MAC-RA scheme is required to serve that purpose.

The BEB scheme is the typical and traditional CSMA/CA mechanism, which was introduced in IEEE 802.11 DCF [18]. A randomly generated backoff value for the contention procedure is used. At the first transmission attempt, the contending STA generates a uniform random backoff value, from the contention window (CW). Initially, CW is set to a minimum value, and after each unsuccessful transmission, its value is doubled until it reaches the maximum defined value. Once a STA successfully transmits its data frame, CW is again reset back to the minimum value. For a network with a heavy load, resetting CW to its minimum value after successful transmission will result in more collisions and poor network performance due to an increase in probability to select similar backoff value for many STAs. Similarly, for fewer contending STAs, the blind exponential increase of CW for collision avoidance causes an unnecessarily long delay due to the wider range for selecting backoff value. Besides, this blind increase/decrease of the backoff window is more inefficient in the highly dense networks proposed for IEEE 802.11ax, because the probability of contention collision increases with the increasing number of STAs. Thus, the current MAC-RA protocol does not allow WLANs to achieve high efficiency in highly dense environments and become a part of future 5G and IoT. Hence, to withstand this challenge, WLAN needs a more efficient and self-scrutinized backoff mechanism to promise enhanced user quality of experience (QoE).

A WLAN system performance can be severely degraded with an increase in the number of contenders, as the collision in the network is directly proportional to the density of the network. This problem statement is assured by the simulation results shown in Figure 2.6. Figure 2.6 plots the number of STAs contending for channel access versus the average channel collision probability in a saturated (always willing to transmit) network environment with CW minimum as 32 and 64. The other simulation parameters are described in Table 5.1. The figure shows that increased network density has a direct relationship with the average channel collision probability; the denser the network, the higher the channel collision probability. In

such a troublesome situation, a more adaptive and self-scrutinized MAC-RA is required by the HEW networks to maintain the performance.

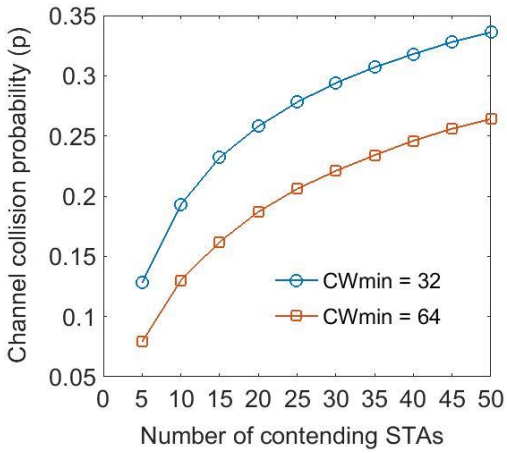

Figure 2-6 Number of contending STAs vs. channel collision probability with CW minimum as 32 and 64.

3. Deep Reinforcement Learning Paradigm

Both academic and industrial communities have recognized that the future smart wireless devices have to rely on enlightened learning and decision-making. Deep learning (DL), as one of the prevailing machine learning (ML) tools, establishes an auspicious paradigm for MAC-RA, as well. As shown in Figure 3.1, we can imagine an intelligent HEW device that is capable of accessing channel resources with the aid of DL techniques. Obviously, an intelligent device learns the performance of a specific action with the objective of preserving a specific performance metric. Later, based on this learning, the intelligent device aims to reliably improve its performance while executing future actions by exploiting previous experience. This chapter highlights the types of DL and their applications in current wireless networks. Later in this chapter, one of the DRL models, Q learning is proposed as an auspicious DRL paradigm for MAC-RA in dense wireless networks.

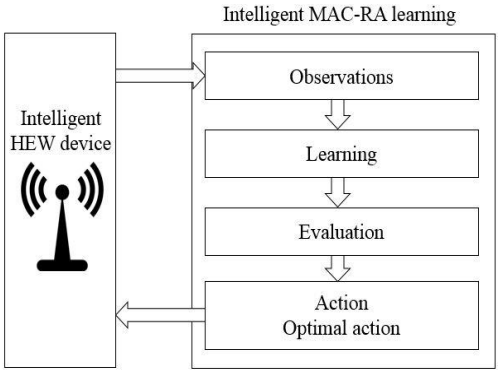

Figure 3-1 Intelligent MAC layer resource allocation (MAC-RA) learning model.

3.1 Deep learning in wireless networks

DL has established comprehensive applications in computer science, such as image processing, audio/video processing, micro-and macro-economic behavioral analysis, and so on [19]. DL algorithms are typically considered supervised or unsupervised learning algorithms. The supervised and unsupervised algorithms specify whether there are categorized samples in the available data (usually known as training data). Recently, another class of DL has emerged, known as dee

reinforcement learning (DRL), which was encouraged by behavioral psychology [20]. In this section, the role of these categories of DL in wireless communication networks is elaborated. Figure 3.2 summarizes the family architecture of DL techniques, models and their potential applications in dense wireless networks.

3.1.1 Supervised deep learning

Supervised DL is learning from a labeled, training dataset supervised by an erudite exterior supervisor. Each labeled training dataset is a depiction of a state composed with a specification, the label, a particular action, and the class to which that particular action belongs. The objective of supervised DL is for the system to infer its retorts so that it acts intelligently in states not present in the labeled training dataset. Although, supervised DL is a significant kind of DL, but without the help of a supervisor and available training dataset, it is not suitable for learning the environment. Thus, for the systems that need to deal interactively, it is often impractical to obtain a sample training dataset of anticipated behavior that is equally precise and descriptive of all the states in which the device has to perform actions in the future. In an unexplored environment, where one would expect DL to be most valuable, a device must be able to learn from its own experience of interaction with the environment [20].

Examples of supervised DL algorithms are regression models, k-nearest neighbor (KNN), the support vector machine (SVM), and Bayesian learning (BL) [21]. Regression analysis (RA) depends on a statistical method for assessing the relations among input parameters. The objective of RA is to envisage the assessment of one or more continuously valued estimation objectives, given the assessment of a vector of input parameters. The estimation objective is a function of the independent parameters. The KNN and SVM techniques are mostly employed to classify/categorize different objects in the system. In the KNN technique, a learner is categorized into a particular category according to the votes of the neighbor agents. The learner is associated with the category that is most common among its k-nearest

neighbors. By contrast, the SVM algorithm uses nonlinear mapping for object classification. First, it converts the original training dataset into a higher measurement where it befits distinguishability. Later, it explores for the optimized linearly separating hyperplane that is accomplished by distinguishing one category of agents from another category of agents [21]. On the other hand, the idea of BL is to estimate a posterior distribution of the target variables, given some inputs and the available training datasets. The hidden Markov model (HMM) is one simple example of reproductive paradigms that can be learned with the help of BL. HMM is a tool considered for expressing probability distributions of the trail of observations in the system. More specifically, it is a generalization method, where the unseen (hidden) variables of the system are associated with each other through a Markov decision process (MDP) [3]. These hidden variables control the particular constituent to be selected for each observation, while being relatively independent of each other.

These examples of supervised DL paradigms can be used for estimating wireless radio parameters that are related to the QoS and QoE requirements of a particular user/device. Like a massive MIMO system of hundreds of radio antennas, the available channel estimation may lead to optimal dimensional search problems, which can easily be learned using any of the above-mentioned supervised learning models. The SVM functions are cooperative for data classification problems. A hierarchical SVM (H-SVM), in which each hierarchical level is comprised of a fixed number of SVM classifiers, was proposed in [22]. H-SVM is used to intelligently estimate the Gaussian channel's noise level in a MIMO system by exploiting the training data. KNN and the SVM can be pragmatic at finding the optimum handover solutions in wireless networks. Similarly, the BL model can be invoked for wireless channel characteristics learning and estimation in future generation ultra-dense wireless networks. For example, Wen et al. estimated both the radio channel parameters in a specific radio cell, as well as those of the intrusive links of the neighboring radio cells, using BL techniques to deal with the pilot contamination

problem faced by massive MIMO systems [23]. Another application of BL was proposed in [24], where a Bayesian inference model was proposed for considering and statistically describing a variety of methods that are proficient at learning the predominant factors for cognitive radio networks (CRNs). Their proposed mechanism covers both the MAC layer and network layer of a wireless network.

3.1.2 Unsupervised deep learning

Unsupervised DL is usually about verdict structure veiled in a collection of the unlabeled training dataset. The expressions supervised DL and unsupervised DL would appear to profoundly categorize most of the DL-based paradigms, but in reality, they are not accurate. The aim of supervised DL is to learn the mapping from an input dataset to an output result where accurate values are given by a supervisor. On the other hand, in unsupervised learning, there is no such external supervisor, and there is only the available input dataset. The objective is to find symmetries in the dataset. There is an edifice of the available dataset space, such that certain patterns occur often, and such patterns can help to understand what action to take in the future for unknown input. In the language of statistics, this is also known as density estimation [21].

Examples of unsupervised DL algorithms are k-means clustering, principle component analysis (PCA), and independent component analysis (ICA). The objective of k-means clustering is to divide user observations into *k* clusters, where each observation is associated with the adjacent cluster. It uses the center of gravity (centroid) of the cluster, which is the mean value of the observation points within that particular cluster. Continuous iteration of the k-means clustering algorithm keeps assigning an agent to the particular cluster in which the centroid is close to the agent based on a similarity metric. This similarity metric is known as Euclidean distance. Later, the in-cluster differences are also minimized until convergence by iteratively updating the cluster centroid [21]. PCA is used to transform a set of possibly associated parameters into a set of unassociated parameters, known as the

principal components (PC). The number of PCs is always less than or equal to the number of original parameters/components. The first PC has the largest possible variance, and each subsequent PC, in turn, has the utmost variance probable under the limitation that it is unassociated with the prior PCs. Basically, the PCs are orthogonal (unassociated), since they are the eigenvectors of the covariance matrix that is symmetric. Unlike PCA, ICA is a statistical method applied to expose unseen elements that inspire sets of haphazard parameters/components within the system [21].

Clustering is one of the common problems in densely deployed wireless networks of 5G and IoT systems, especially in heterogeneous network environments with diverse cell sizes. In such cases, the small cells have to be wisely grouped to avoid interference using coordinated multi-point transmission, whereas the mobile devices are grouped to follow an optimum offloading strategy. The devices are grouped in device-to-device (D2D) wireless networks to attain high energy-efficiency, and the WLAN users are grouped to uphold an optimum access point (AP) association. Xia *et. al.* proposed a hybrid scenario in order to diminish inclusive wireless traffic by encouraging the exploitation of a high-capacity optical infrastructure [25]. They formulated a mixed-integer programming (MIP) problem to cooperatively optimize both network gateway splitting and the virtual radio channel provision based on typical k-means clustering. Both PCA and ICA are formulated to recover statistically autonomous source signals from their linear combinations using powerful statistical signal processing techniques. One of their key applications is in the area of intrusion-detection issues in WLANs, which depends on traffic monitoring. Besides, similar issues may also be resolved in the ultra-dense wireless communications technologies of 5G and IoT systems. PCA and ICA can also be invoked to classify user behavior in CRNs. In [26], the authors applied PCA and ICA in a smart grid scenario of IoT systems to improve the concurrent wireless transmissions of smart devices set up in the smart home. The statistical possessions of the received signals were oppressed to blindly isolate them

using ICA. Their proposed mechanism enhances transmission capability by evading radio channel assessment as well as data security by excluding any wideband intrusion.

3.1.3 Deep reinforcement learning

Deep reinforcement learning (DRL) is motivated by behaviorist sensibility and a control philosophy, where a learner can achieve its objective by interacting with and learning from its surroundings. In DRL, the learner does not have clear information on whether it has come close to its target objective. However, the learner can observe the environment to augment the aggregate reward in an MDP [27]. DRL is one DL technique that learns about the environment, what to do, and how to outline circumstances to current actions in order to maximize a numerical reward signal. Mostly, the learner is not informed about which actions to take, yet it has to learn which actions produce the maximum reward by trying them. In the utmost exciting and inspiring situations, it is possible that actions will affect not only the instant reward but also the following state, and through that, all succeeding rewards. MDPs offer a precise framework for modeling decision making in particular circumstances, where the consequences are comparatively haphazard, and the decision maker partially governs the consequences.

Examples of DRL are partially observable MDP (POMDP) and Q learning (QL). POMDP might be seen as speculation with MDP, where the learner is inadequate to straightforwardly perceive the original state transitions, and thus, only has constrained information. The learner has to retain the trajectory of the probability distribution of the appropriate states, based on a set of annotations, as well as the probability distribution of both the observation probabilities and the original MDP [21]. QL might be conjured up to discover an optimum strategy for taking action from any finite MDP, particularly when the environment is unknown. POMDP is an RL technique that does not follow a model, and it can also be combined with MDP models. In such a case, the QL paradigm also covers a set of states where an

agent can make a decision on an action from a set of available actions. By performing an action in a particular state, the agent collects a reward, with the objective being to exploit its collective rewards. A collective reward is illustrated as a Q-function, and is updated in an iterative approach after the agent carries out an action and attains the subsequent reward [21].

The uses of POMDP paradigms create vital tools for supportive decision making in IoT systems, where the IoT devices may be considered learners, and the wireless network constitutes the environment. In a POMDP problem, the technique first postulates the environment's state space and the learner's action space, as well as indorsing the Markov property among the states. Secondly, it constructs the state transition probabilities formulated as the probability of navigating from one state to another under a specific action. The third and final step is to enumerate both the learer's instant reward and its long-term reward via Bellman's equation [20]. Later, a wisely constructed iterative algorithm may be considered to classify the optimum action in each state. The applications of POMDP comprise the network selection problems of heterogeneous networks, channel sensing, and user access in CRNs. In [28], the authors proposed a mechanism for transmission power control problems of energy-harvesting systems, which were scrutinized with the help of the POMDP model. In their proposed investigation, the battery state, the channel state, and data transmission and data reception states are defined as the state space, and an action by the agent is related to transmitting a packet at a certain transmission power. QL, usually in aggregation with the MDP models, has also been used in applications of heterogeneous networks. Alnwaimi et al. presented a heterogeneous, fully distributed, multi-objective strategy for optimization of femtocells based on a QL model [29]. Their proposed model solves both the channel resource allocation and interference coordination issues in the downlink of heterogeneous femtocell networks. Their proposed model acquires channel distribution awareness, and classifies the accessibility of vacant radio channel slots for the establishment of

opportunistic access. Later, it helps to pick sub-channels from the vacant spectrum pool.

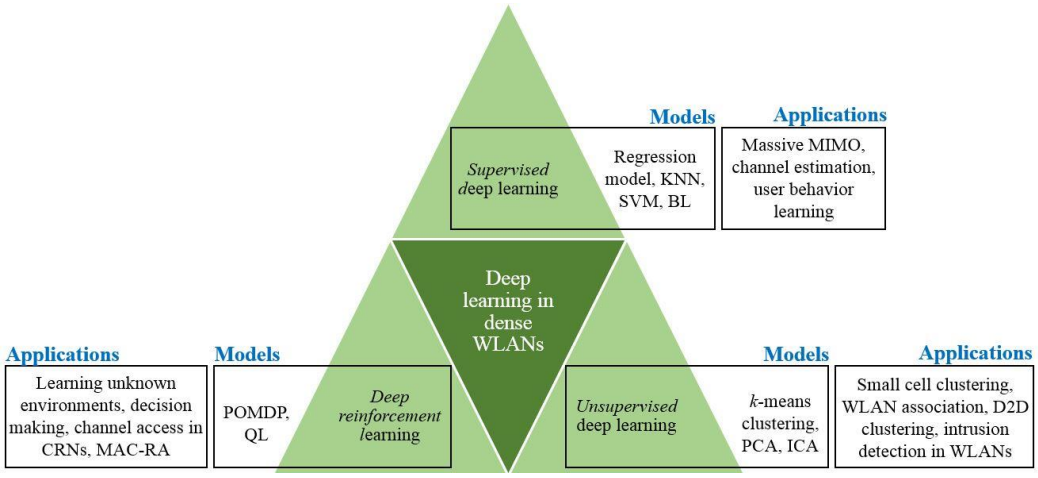

Figure 3-2 Deep learning family architecture: models and their potential applications in dense WLANs.

3.2 Q learning as a MAC-RA paradigm

As described in the previous section, QL has already been extensively applied in heterogeneous wireless networks. In this section, the main functional structure of the QL algorithm is described, and in a later subsection, the use of QL as a future paradigm for the backoff mechanism in a distributed coordination function (DCF) is suggested for dense wireless networks.

3.2.1 Q Learning Algorithm

The QL algorithm utilizes a form of DRL to solve MDPs without possessing complete information. Aside from the learner and the environment, a QL system has four main sub-elements: a policy, a reward, a Q-value function, and sometimes a model of the environment as an optional entity (most of the QL algorithms are model-free), as shown in Figure 3.3.

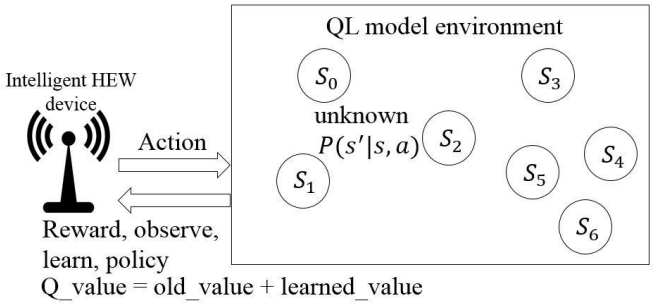

Figure 3-3 Q-learning model environment for an intelligent HEW

Policy

The learner's way of behaving at a particular time is defined as a policy. It resembles what in psychology would be called a set of stimulus–response associations. In some circumstances, the policy can be a modest utility or a lookup table; in others, however, it may comprise extensive computations, like an exploration process. The policy is fundamental for a QL learner, in the sense that it alone is adequate to determine behavior. Generally, policies might be stochastic. A policy decides which action to take in which state.

Reward

A reward expresses the objective in a QL problem. In each time step, the environment passes to the QL learner a particular quantity called the reward. The learner's exclusive goal is to exploit the total reward it obtains over the long run. The reward describes the pleasant and unpleasant events for the learner. Reward signals are the instant and crucial topographies of the problem faced by the learner. The reward signal is the key basis for changing the policy. For example, if the current action taken by a policy is followed by a low reward, a learner may decide to select other actions in future.

Q-value function

Though the reward specifies what is good at one instant, a value function stipulates what is good in the end. Thus, the value (known as the Q-value) of a state

is the aggregate amount of rewards a learner can presume to accumulate over the future, starting from that state. For example, a state might continuously produce a low instant reward, but still have a high Q-value because it is repeatedly trailed by other states that produce high rewards. To make a WLAN environment correspondent, rewards are somewhat like a high channel collision probability (unpleased) and a low channel collision probability (pleased), whereas Q-values resemble a more sophisticated and prophetic verdict of how pleased or unpleased the learner (STA) is in a particular state (e.g. the backoff stage). If there is no reward, there will be no Q-value, and the only purpose for estimating the Q-value is to attain additional rewards. It is the Q-value that a learner is the most anxious about when making and assessing verdicts. A learner selects optimum actions based on Q-value findings. It seeks actions that carry states of a maximum Q-value, not a maximum reward, because these actions attain the highest amount from the rewards for the learner over the long run.

Environment model

An optional element of QL is a model of the system, which somewhat mimics the performance of the environment. Typically, it allows inferences to be made about how the environment will perform. For example, given a state and an action, the model might envision the subsequent state and the next reward. Environment models are used for planning a way to decide on a sequence of actions by considering latent future situations. In an example of a WLAN system, a device (the learner) would like to plan its future decisions based on a given state (e.g. the backoff stage) and action, along with its rewards (e.g. channel collision probability).

3.2.2 Scope and Limitations of QL

As discussed above, QL depends strongly on the notion of the state as input to the policy and the Q-value function. Informally, we can think of the state as a flag passing to the learner with some sense of how the environment is at a specific time. A large portion of QL techniques are organized around evaluating Q-value functions;

however, it is not entirely essential to do this to take care of DRL problems. For instance, approaches like genetic algorithms, genetic programming, simulated forging, and other optimization algorithms have been utilized to approach DRL problems while never engaging value functions [30]. These evolutionary approaches assess the lifetime conduct of numerous non-learners, each utilizing an alternate policy for interfacing with the environment and selecting those actions that are able to acquire the most rewards. If the space of policies is adequately small, or can be organized so that the best policies are common or simple to discover, or if a considerable measure of time is available for the search, then evolutionary approaches can be viable. Furthermore, evolutionary approaches have focal points for problems in which the learner cannot detect the entire state of the environment. In contrast to evolutionary approaches, QL techniques learn while interfering with the environment. Techniques ready to exploit the details of individual behavioral interactions can be substantially more productive than evolutionary strategies in many types of wireless network.

4. *intelligent* Q learning-based Resource Allocation (*i*QRA)

The QL-based MAC-RA scheme can be used to guide future densely deployed WLAN devices and to allocate channel resources more efficiently. When a WLAN device is deployed in a new environment, usually no data are available on historical scenarios. Therefore, QL algorithms are the best choice to observe and learn the environment for optimal policy selection. In a densely deployed WLAN, channel collision is the most vital issue causing performance degradation. Since QL finds solutions through the experience of interacting and learning with an environment, it is proposed using it to model the optimal contention window in MAC-RA. In other words, a learner (STA) controls the CW selection intelligently with the aid of the QL-based algorithm. In a DCF-based backoff mechanism, STAs can be equipped with the intelligent QL algorithm. The policy is the decision of an STA to change the CW size (that is, to take an action to move to the next backoff stage and increase the CW size, or to move to the previous/first backoff stage and decrease/reset the CW size, or possibly stay at the same backoff stage with no change to the CW). The reward function captures the gain of each action performed in any state. For example, a reward can be channel collision probability, channel access delay, or the packet loss ratio (PLR) experienced by the STA at a specific state (backoff stage).

Next subsection elaborates a channel observation-based scaled backoff (COSB) protocol to tackle the blindness problem of conventional BEB algorithm. Later, in the second subsection, an *intelligent* QL-based resource allocation (*i*QRA) is proposed to optimize the performance of COSB.

4.1 Channel observation-based scaled backoff (COSB) mechanism

In the proposed COSB protocol, after the communication medium has been idle for a DIFS period, all the STAs competing for a channel proceed to the backoff procedure by selecting a random backoff value B as shown in Figure 4.1. The time immediately following an idle DIFS is slotted into observation time slots (γ). The

duration of a γ slot is either a constant slot time σ during an idle period or a variable busy (successful or collided transmission) period. While the channel is sensed to be idle during σ, B decrements by one. A data frame is transmitted after B reaches zero. In addition, if the medium is sensed to be busy, the STA freezes B and continues sensing the channel. If the channel is again sensed to be idle for DIFS, B is resumed. Each individual STA can proficiently measure channel observation-based conditional collision probability p_{obs}, which is defined as the probability that a data frame transmitted by a tagged STA fails. The time is discretized in B_{obs} observation time slots, where the value of B_{obs} is the total number of γ observation slots between two consecutive backoff stages as shown in Figure 4.1. A tagged STA updates p_{obs} from B_{obs} of backoff stage b_i at the ith transmission as,

$$p_{obs} = \frac{1}{B_{obs}} \times \sum_{k=0}^{B_{obs}-1} S_k, \tag{4.1}$$

where for an observation time slot k, $S_k = 0$ if γ is empty (idle) or the tagged STA transmits successfully, while $S_k = 1$ if γ is busy or the tagged STA experiences collision as shown in Figure 4.1. In the figure, STA 1 randomly selects its backoff value $B = 9$ for its b_i backoff stage. Since STA 1 observes nine idle slot times, two busy periods, and one collision ($B_{obs} = 9 + 2 + 1 = 12$), p_{obs} is updated as $\frac{2+1}{B_{obs}} = \frac{3}{12} = 0.25$ in the next backoff stage b_{i+1}.

According to the channel observation-based conditional collision probability p_{obs}, the adaptively scaled contention window value is $W_{b_{i+1}}$ at backoff stage b_{i+1} of the transmission time $i + 1$, where $b_{i+1} \in (0, m)$ for the maximum m number of backoff stages, and i is the discretized time for the data frame transmissions of a tagged STA. More specifically, when a transmitted data frame has collided, the current contention window $W_{b_i}$of backoff stage b_i at the ith transmission time slot is scaled-up according to the observed p_{obs} at the ith transmission, and when a data frame is transmitted successfully, the current contention window $W_{b_i}$is scaled-down according to the observed p_{obs} at the ith transmission. Unlike the BEB (where

backoff stage is incremented for each retransmission and resets to zero for new transmission as shown in Figure 4.2(a)), the backoff stage b_i in COSB at the ith transmission has the following property of increment or decrement:

$$b_{i+1} = \begin{cases} min[b_i + 1, m], \text{collision at i}^{\text{th}} \text{ transmit} \\ max[b_i - 1, 0], \text{success at i}^{\text{th}} \text{transmit} \end{cases}. \tag{4.2}$$

Figure 4.2(b) shows that the backoff stage in COSB does not reset after a successful transmission. Since the current backoff stage represents the number of collisions or successful transmissions of a tagged STA, it helps to scale the size of CW efficiently. The incremented or decremented backoff stage b_i results in scaling-up or scaling-down of the current contention window, respectively. The scaling-up and scaling-down of the contention window operates as follows:

$$W_{b_{i+1}} = \begin{cases} min[2^{b_{i+1}} \times W_{min} \times \omega^{p_{obs}}, W_{max}], \text{collision at i}^{\text{th}} \text{ transmit} \\ max[2^{b_{i+1}} \times W_{min} \times \omega^{p_{obs}}, W_{min}], \text{success at i}^{\text{th}} \text{ transmit} \end{cases}, \tag{4.3}$$

where ω is a constant design parameter to control the optimal size of the contention window and is expressed as $\omega = W_{min}$. The W_{min}, and W_{max} are the minimum CW and maximum CW values.

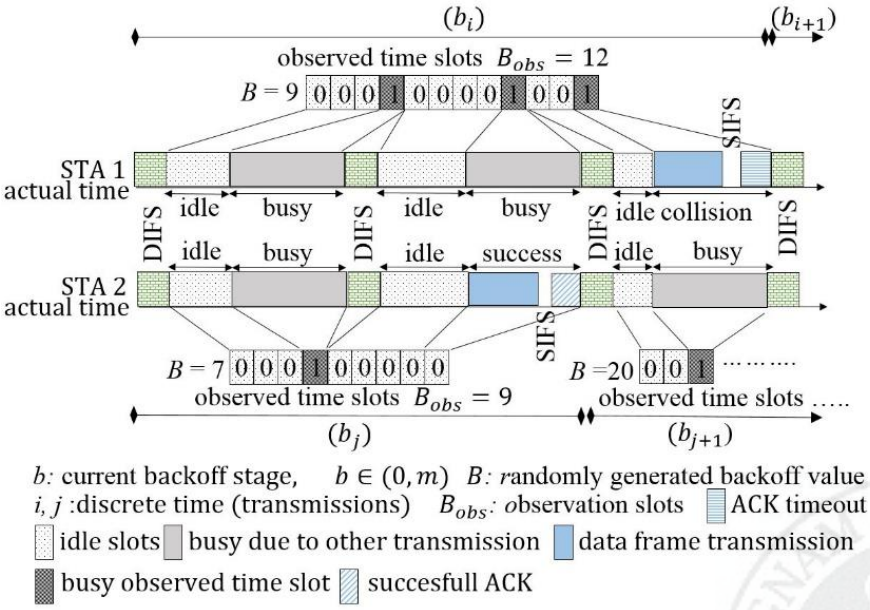

Figure 4-1 Channel observation mechanism of channel observation-based scaled backoff (COSB) during the backoff procedure.

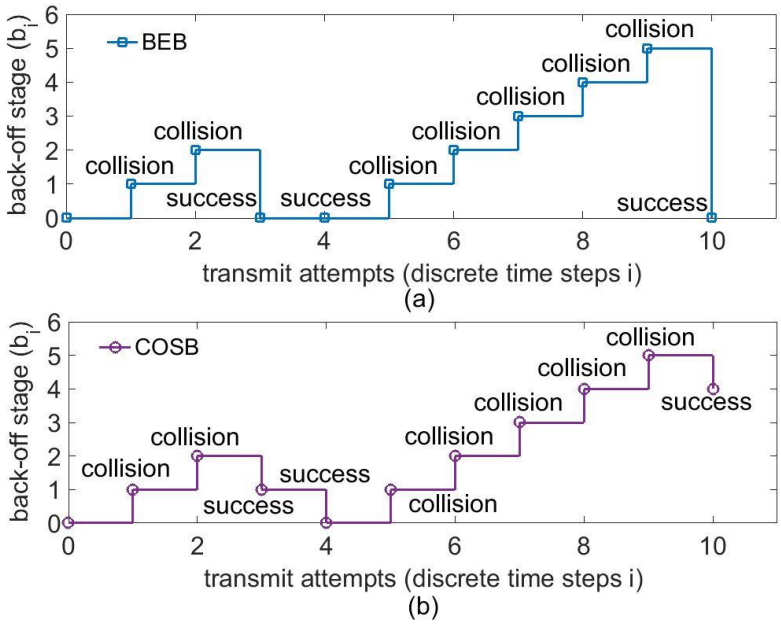

Figure 4-2 Backoff stage after collision/successful transmission; (a) backoff stage increment/reset in binary exponential backoff (BEB); and (b) backoff stage increment/decrement in COSB.

4.1.1 Analytical modeling of COSB

An analytical model of COSB is formulated to mathematically affirm the validity of the mechanism. The ideal channel conditions, that is no hidden STA and capture effects, are assumed for this purpose. The fixed number of STAs are assumed, each of them is always willing to transmit the data frame, i.e., the network is assumed as saturated traffic environment. Initially, the behavior of a tagged STA is studied with a discrete-time Markov chain model (DTMC) [31] [32], and the stationary transmission probability τ is obtained for a tagged STA. Since proposed COSB does not reset backoff stage to its initial value (that is to zero) after successful transmission, the transmission attempt for every new data frame remains recursive within the backoff stage state dimension. To accurately analyze the performance of COSB, a recursive discrete-time Markov chain model (R-DTMC) is formulated. Later, by knowing the exact events that can occur on the communication channel within a randomly selected slot-time, the normalized throughput and channel access delay of the proposed COSB mechanism can be formulated.

Recursive Discrete-Time Markove Chain (R-DTMC) Model

Consider there are n number of STAs competing for the channel in a WLAN. In the saturated condition, each STA has immediately a data frame available for transmission after each successful transmission. Thus, due to the consecutive data frame transmission, each data frame needs to wait for a random backoff time before transmitting. Let b be the backoff stage counter for a tagged STA and m be the maximum number of backoff stages can be experienced for a data frame, that is $b \in (0, m)$, such that $W_b = 2^b \times W_{min} \times \omega^{p_{obs}}$ for b^{th} backoff stage and $W_{max} = 2^m \times W_{min} \times \omega^{p_{obs}}$ for the mth backoff stage contention window, where W_b is the contention window size at b^{th} backoff stage and p_{obs} is the observed channel collision probability. Let us adopt the notation $W_{b+1} = 2^{b+1} \times W_{min} \times \omega^{p_{obs}}$, for the adaptively scaled-up contention window for $b + 1$ backoff stage, when transmission is failed at the bth backoff stage. Similarly, let $W_{b-1} = 2^{b-1} \times W_{min} \times \omega^{p_{obs}}$ be the adaptively scaled-down contention window for $b - 1$ backoff stage, when successfully transmitted at the bth backoff stage.

Assume $\Omega(t)$ be the function for stochastic process representing the backoff counter u for a tagged STA, where $u \in (0, W_{cur} - 1)$. Since time is discretized as an integer time scale, t and $t + 1$ correspond to the beginning of two consecutive transmission time slots, and the backoff time counter of each STA decrements at the beginning of each slot time. Figure 4.1 illustrates that the backoff time decrements when the communication channel is sensed as idle (σ), and it stops when the channel is sensed busy, which may be due to a successful or unsuccessful transmission of any other STA. Therefore, the time interval between two consecutive slot time beginnings may be much longer and different from the idle slot time size, i.e. σ. Let $\pi(t)$ be the stochastic process representing the backoff stages of a tagged STA. The key articulation in R-DTMC model is that, regardless of the number of retransmission attempts at each data frame transmission attempt, each data frame collides with a practically observed and independent collision probability p_{obs}.

With these assumptions, COSB can be modeled as the two dimensional Markov process $\{\pi(t),\ \Omega(t)\}$ with the R-DTMC as depicted in the Figure 4.3.

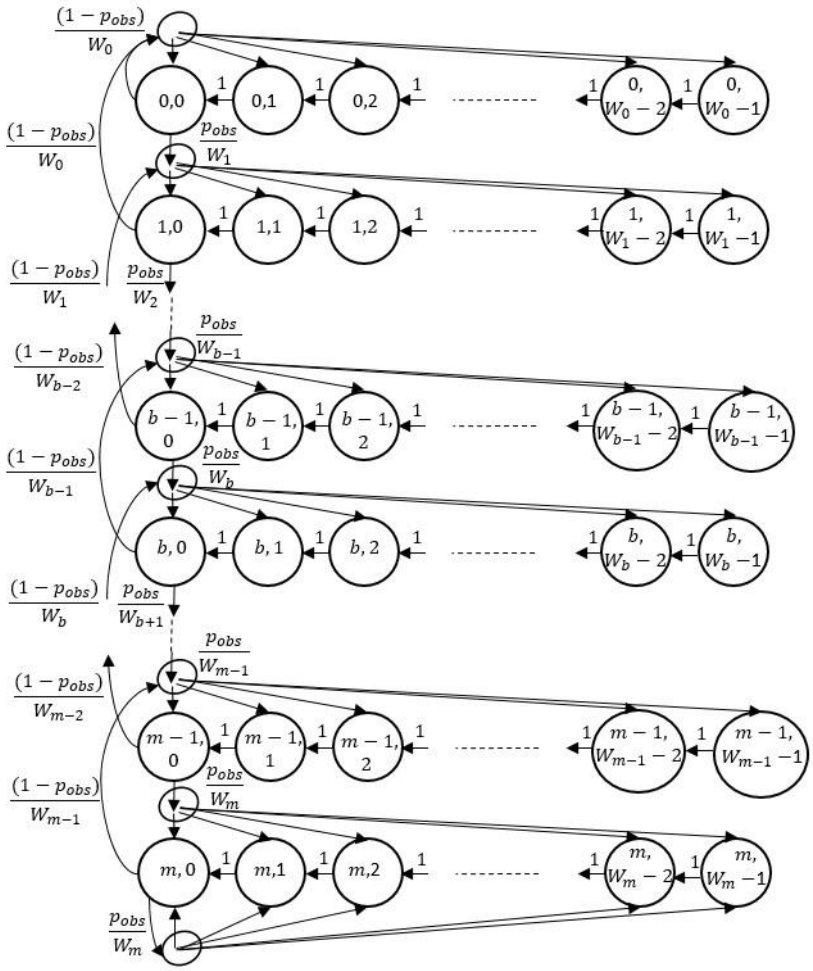

Figure 4-3 Recursive discrete-time Markov chain model (R-DTMC) for channel observation-based scaled back-off (COSB) mechanism.

In this R-DTMC, the transition probabilities are described as follows,

1. The tagged STA remains at the first backoff stage after a successful transmission on first backoff stage with the probability,

$$P\{(0,u)|(0,0)\} = (1-p_{obs})/W_0,\ \ u \in (0, W_0-1) \tag{4.4}$$

2. The backoff counter decrements when the channel is sensed as idle with the probability,

$$P\{(b,u)|(b,u+1)\} = 1,\ \ b \in (0,m), u \in (0, W_b-2) \tag{4.5}$$

3. The tagged STA scales-up the current contention window and moves to the next stage b if a data frame transmission failed on backoff stage $b-1$ with the probability,

$$P\{(b,u)|(b-1,0)\} = p_{obs}/W_b, \quad b \in (0,m), u \in (0, W_b - 1) \tag{4.6}$$

4. The tagged STA scales-down the current contention window and decrements its backoff stage for next transmission attempt to $b-1$ after a successful transmission on backoff stage b with the probability,

$$P\{(b-1,u)|(b,0)\} = (1-p_{obs})/W_{b-1}, \quad b \in (0,m), u \in (0, W_b - 1) \tag{4.7}$$

5. The tagged STA remains at the mth backoff stage after an unsuccessful transmission with the probability,

$$P\{(m,u)|(m,0)\} = p_{obs}/W_m, \quad u \in (0, W_m - 1) \tag{4.8}$$

In particular, to the above transition probabilities, as considered in the equation (4.6), when a data frame transmission is collided at backoff stage $b-1$, the backoff stage increases to b, and the new backoff value is uniformly chosen from the adaptively scaled-up contention window W_b. On the other hand, equation (4.7) describes that when a data frame transmission is successful at backoff stage b, the backoff stage decreases to $b-1$, and the new backoff value is uniformly chosen from the adaptively scaled-down contention window W_{b-1}. In case the backoff stage reaches the value m (that is the maximum backoff value), it is not increased in subsequent data frame transmission attempt. Let us assume that $d_{b,u} = \lim_{t\to\infty} P\{\pi(t) = b, \Omega(t) = u\}$, $b \in (0,m), u \in (0, W_b - 1)$ be the stationary distribution of the R-DTMC. From Figure 4.3, each state transition probability can be written as,

$$d_{1,0} = \frac{p_{obs}}{1-p_{obs}} d_{0,0}. \tag{4.9}$$

If $\rho = \frac{p_{obs}}{1-p_{obs}}$, the above equation can be written as, $d_{1,0} = \rho \times d_{0,0}$. similarly, $d_{b,0} = \rho \times d_{b-1,0}$ where $d_{b-1,0} = \rho \times d_{b-2,0}$ till $d_{1,0} = \rho \times d_{0,0.}$ Therefore, we can write,

$$d_{b,0} = \rho^b \times d_{0,0}, \qquad 0 < b < m. \tag{4.10}$$

Now for the backoff stage m, the $d_{m,0}$ can be written as,

$$d_{m,0} = \rho^m \times d_{0,0} \tag{4.11}$$

Owing to the Markov process-based chain regularities, for each $u \in (1, W_b - 1)$, the stationary distribution for $\{\pi(t), \Omega(t)\}$ can be written as,

$$d_{b,u} = \frac{W_b - u}{W_b} \times \tag{4.12}$$

$$\begin{cases} (1 - p_{obs}) \times d_{b+1,0} & b = 0 \\ p_{obs} \times d_{b-1,0} + (1 - p_{obs}) \times d_{b+1,0} & 0 < b < m. \\ p_{obs} \times \left(d_{m-1,0} + d_{m,0}\right) & b = m \end{cases}$$

The recursive characteristic of state transition probabilities can be combined as,

$$\sum_{b=0}^{m-1} (b+1) \times d_{b,0} + m \times d_{m,0} = d_{0,0} \left(\frac{1 - \rho^m - m\rho^m(1-\rho)}{(1-\rho)^2}\right). \tag{4.13}$$

From (4.9-4.11) and (4.13), (4.12) can be re-written as,

$$d_{b,u} = \frac{(W_b - u)(b+1)}{W_b} d_{b,0}, \qquad b \in (0, m), u \in (0, W_b - 1) \tag{4.14}$$

From (4.9 - 4.11) and (4.14), all the values $d_{b,u}$ are expressed as function of $d_{0,0}$ and channel observation-based practical conditional collision probability p_{obs}. The $d_{0,0}$ is finally determined by normalizing the R-DTMC states as follows,

$$1 = \sum_{b=0}^{m-1} (b+1) \times d_{b,0} \sum_{u=0}^{W_b - 1} \frac{W_b - u}{W_b} + \frac{m \times (W_m - u)}{W_m} d_{m,0}, \tag{4.15}$$

From $W^* = W_{min} \times \omega^{p_{obs}}$ and after few mathematical steps, above normalization relation can be written as,

$$1 = \frac{d_{0,0}}{2} \left[W^* \left(\frac{1 - (2\rho)^m - m(2\rho)^m(1 - 2\rho)}{(1 - 2\rho)^2} \right) + \left(\frac{1 - \rho^m - m\rho^m(1 - \rho)}{(1 - \rho)^2} \right) \right] \tag{4.16}$$

Finally, we get $d_{0,0}$ as follows,

$$d_{0,0} = \frac{2(1-\rho)^2(1-\rho^m)}{\left(1-\rho^m - m\rho^m(1-\rho)\right)} \times \frac{1}{\left((1-2\rho)(1-\rho^m)(W^*+1)+\rho W^*(1-\rho)(1-(2\rho)^m)\right)}. \tag{4.17}$$

Since, a transmission occurs only when the backoff counter of the STA reaches zero regardless of the backoff stage, transmission probability τ can be expressed as follows,

$$\tau = \sum_{b=0}^{m-1}(b+1)d_{b,0} + md_{m,0} = d_{0,0}\left(\frac{1-\rho^m - m\rho^m(1-\rho)}{(1-\rho)^2}\right). \tag{4.18}$$

Further, after performing few mathematical steps to (4.18) using the value of $d_{0,0}$ from (4.17) we get,

$$\tau = \frac{2}{\left(W^* + \rho W^*\left(\frac{\sum_{b=0}^{m-1}(2\rho)^b}{\sum_{b=0}^{m-1}(\rho)^b}\right)+1\right)}. \tag{4.19}$$

However, in general τ depends on the practical collision probability p_{obs}, which is always unknown until the channel is observed for the busy slots. A transmitted data frame encounters the collision if at least one of the $n-1$ remaining STAs transmit. Since each of the transmission in the system sees this collision in the same state, a steady state can easily be yielded as [32],

$$p_{obs} = 1-(1-\tau)^{n-1}. \tag{4.20}$$

These two (τ and p_{obs}) are monotonic non-linear systems which can be numerically solved for each other.

Normalized Throughput

Let θ be the normalized throughput of the network and be defined as the fraction of the communication channel used for successful transmission of the data payload. To compute θ, let τ_{tr} be the probability that there is at least one transmission in the considered slot time. Since there are n number of STAs in the system contending

for the medium and each transmits with probability τ, the transmission probability τ_{tr} can be defined as,

$$\tau_{tr} = 1 - (1 - \tau)^n. \tag{4.21}$$

If the probability τ_s that a transmission is successful is given by the probability that only one STA transmits in the considered slot time, τ_s can be obtained as,

$$\tau_s = \frac{n\tau(1-\tau)^{n-1}}{\tau_{tr}} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n}. \tag{4.22}$$

Thus, θ can be expressed as the ratio,

$$\theta = \frac{E[mean\ payload\ transmitted\ in\ a\ slot\ time]}{E[total\ lenght\ of\ a\ slot\ time]}. \tag{4.23}$$

Assume $E[P]$ is the average data frame payload size (assuming that all the data frames have the same fixed size), then the slot time for transmitting average payload data successfully can be obtained as, $\tau_{tr}\tau_s E[P]$, since $\tau_{tr}\tau_s$ is the probability for the successful transmission of a data frame in a given slot time. The average length of a given slot time is the sum of three cases; no transmission in a slot time that is $(1-\tau_{tr})\sigma$, a successfully transmitted data frame that is $\tau_{tr}\tau_s$, and a collision that is $\tau_{tr}(1-\tau_s)$. Finally, the relation (4.23) can be written as follows:

$$\theta = \frac{\tau_{tr}\tau_s E[P]}{(1-\tau_{tr}).\sigma + \tau_{tr}\tau_s.T_s + \tau_{tr}(1-\tau_s).T_c} \tag{4.24}$$

where T_s and T_c are the average time the communication channel has been busy due to successful transmission and collision, respectively. For analytical evaluation, the values of $E[P]$, T_s, T_c, and idle slot time σ must be expressed with the same time unit. Let $P_{hdr} = PHY_{hdr} + MAC_{hdr}$ be the time to transmit a data frame header, and δ be the channel propagation delay. If ACK is the time to receive an acknowledgement, T_s and T_c can be obtained as,

$$T_s = P_{hdr} + E[P] + SIFS + \delta + ACK + DIFS + \delta, \tag{4.25}$$

$$T_c = P_{hdr} + E[P] + DIFS + \delta. \tag{4.26}$$

The corresponding values for T_s and T_c depend upon the 802.11 standard. The PHY and MAC layer parameters to compute T_s and T_c are shown in the Table 5.1.

4.2 *i*QRA algorithm

The proposed *i*QRA mechanism for performance optimization of COSB, consists of a set of states S (backoff stages), where an intelligent HEW device performs an action a (increments/decrements) according to COSB mechanism. By performing action a following a policy Φ in a particular state s, the device collects a reward r, that is $r(s,a)$ with the objective to exploit the collective reward, $Q(s,a)$ which is a Q-value function. Figure 4.4 depicts the model environment with its elements for the proposed *i*QRA mechanism.

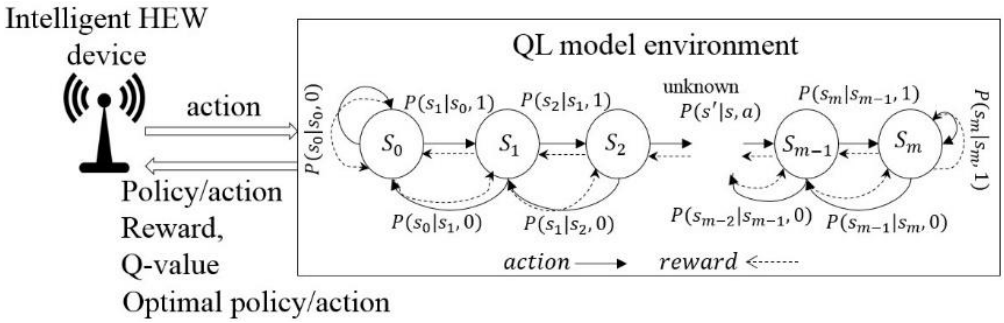

Figure 4-4 *intelligent* Q learning–based resource allocation (*i*QRA); the system environment and its elements.

Let $S = \{0,1,2,\dots,m\}$ denotes a finite set of m possible states of the HEW environment, and let $A = \{0,1\}$ represents a finite set of permissible actions to be taken, where zero indicates decrement, and 1 indicates increment (As described earlier, in COSB, there are two possible actions: increase or decrease of the backoff stage). At time slot t, the STA observes the current state (s), i.e. $s_t = s \in S$, and takes an action (a), i.e. $a_t = a \in A$ based on policy Φ. As mentioned before, the default policy of a device in COSB is to increment its state if collision happened and decrement for successful transmission. Thus, action a_t changes the environmental state from s_t to $s_{t+1} = s' \in S$ according to $\Phi(a|s) = \begin{cases} s' = s+1\text{, if collision} \\ s' = s-1\text{, if succesful} \end{cases}$. The objective of the QL algorithm is to discover an optimal policy Φ^{opt} that exploits the total expected reward (optimal Q-value), which is given by a Bellman's equation [20]:

$$Q^{opt}(s,a) = \mathbb{E}\{r_t(s_t,a_t) + \beta \times max_{a'} Q^{opt}(s',a') | s_t = s, a_t = a\} \quad (4.27)$$

Since the reward may easily get unbounded, a discounted reward factor, β ($0 < \beta < 1$), is used. In the QL algorithm, $Q(s,a)$ estimates the reward as the cumulative reward and is updated as follows:

$$Q(s,a) = (1-\alpha) \times Q(s,a) + \alpha \times \Delta Q(s,a) \quad (4.28)$$

where α is the learning rate and is defined as $0 < \alpha < 1$. The learning occurs quickly, based on the improved learning estimate $\Delta Q(s,a)$, and is expressed as

$$\Delta Q(s,a) = \{r(s,a) + \beta \times max_{a'} Q(s',a')\} - Q(s,a) \quad (4.29)$$

The $max_{a'} Q(s',a')$ defines the best-estimated value for the prospective state s'. In the long run, $Q(s,a)$ converges to the optimal Q-value $Q^{opt}(s,a)$, that is, $\lim_{t\to\infty} Q(s,a) = Q^{opt}(s,a)$. The naivest policy for action selection can be to pick one of the actions with the maximum measured Q-value (that is, exploitation). This exploitation method follows the optimal policy Φ^{opt} and is known as a greedy action ($a^{\Phi^{opt}}$) selection method which can be written as,

$$a^{\Phi^{opt}} = \Phi^{opt}(a|s) = argmax_a \, Q^{opt}(s,a) \quad (4.30)$$

where $argmax_a$ in above equation signifies $a^{\Phi^{opt}}$, for which the expression that follows is exploited. The immediate reward is maximized by continuous exploitation of the greedy action selection method. A simple substitute is to exploit most of the time, but every once in a while, the STA explores all the permissible actions independent of $a^{\Phi^{opt}}$, that is with default policy Φ (known as exploration) with probability ε. The greedy and non-greedy selection of actions is known as the ε-greedy method [20]. The main feature of the ε-greedy technique is that, as the number of instances increases, every action guarantees the convergence of learning estimate $\Delta Q(s,a)$. In HEW, for dense WLANs, the STA would exploit to improve throughput performance, and would explore to know the dynamicity of the WLAN environment.

Since COSB conducts p_{obs} at every transmission attempt (state), we can considered p_{obs} as reward of the action. Therefore, p_{obs} from equation (4.1) is expressed as the reward in order to minimize channel collision probability. The reward given by action a_t taken in state s_t at time t is described as,

$$r_t(s_t, a_t) = 1 - p_{obs} \tag{4.31}$$

The above statement indicates how pleased the STA was with its action in state s_t. In Figure 4.4, the STA moves from one state to another state with $1 - p_{obs}$ as a reward. The STA observes and learns the environment to optimize the backoff process. Algorithm 1 depicts the steps performed by the proposed *i*QRA mechanism to optimize the COSB protocol.

Algorithm 1. COSB performance optimization using *i*QRA

```
GLOBAL: Initialize r(s,a), and Q(s,a).
Function: Select CW using iQRA
        Input: p_obs
       Output: Optimized CW
  Initialize: cur_rew = 0, ΔQ(s,a) = 0
  Calculate reward according to equation (4.31)
  Update reward matrix for r(s,a) = cur_rew
  Calculate improved learning estimate ΔQ(s,a) according to equation
  (4.29)
  Update Q(s,a) according to equation (4.28)
  Pick a random value to explore or exploit (ε-greedy method)
  if (exploit)
     Find a^{Φ^opt} according to equation (4.30)
     Scale CW according to the optimal action
  else (explore)
     Scale CW using COSB mechanism
  end if
  return CW
end Function
```

4.2.1 Computational complexity

The computational complexity of the proposed *i*QRA mechanism is based on the learning phase and is independent of the network size. The STA learns the system

by exploring different permissible actions in every specific state using COSB mechanism. However, as soon as the environment is learned, the best action can be exploited in any given state in a ε-greedy manner, resulting in the optimal solution. Since *i*QRA performs only a fixed number of computations (a fixed number of actions and states), its computational complexity per iteration/episode can be written either as $O(1)$ if explores (that is the computational complexity of COSB), or as $O(mln(i))$ for $i \in (1, m)$ of m number of states if exploits. The best case for the computational complexity is when there is only one possible state to move at any state, that is $m = 1$, and the worst case arises with the m number of states. The computational complexity of *i*QRA mechanism is checked for $m = 6$, which is a default value of number of backoff stages in most of the IEEE 802.11 standards (other simulation parameters are given in Table 5.1). The average processing time obtained for BEB, COSB and *i*QRA are $0.279\mu s$, $0.956\mu s$ and $3.44\mu s$, respectively as shown in Figure 4.5. The increase in computational/processing time for *i*QRA is obvious due to additional inference functionalities as depicted by Figure 4.6.

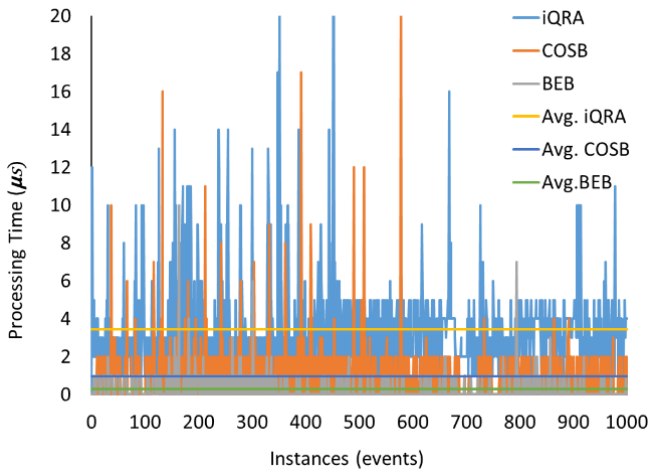

Figure 4-5 Complexity time comparison of BEB, COSB and *i*QRA for 1000 instances (events).

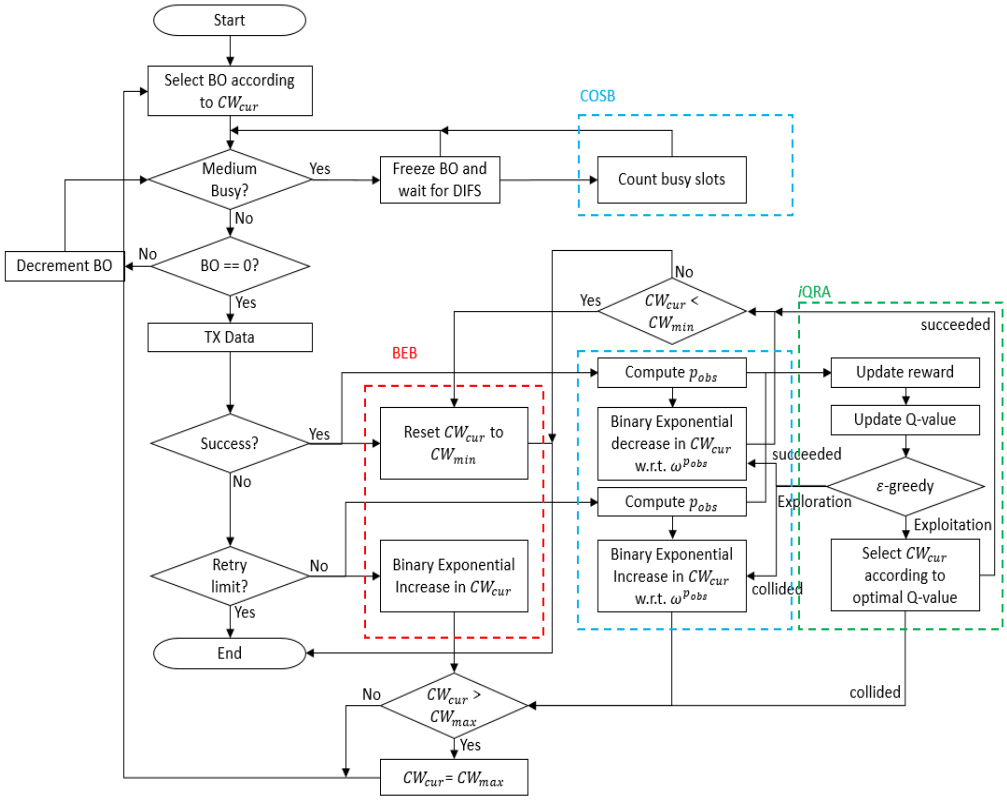

Figure 4-6 Functional comparison of BEB, COSB and *i*QRA

5. Performance Evaluation

5.1 Simulation scenarios and parameters selection

The proposed learning-based *i*QRA mechanism is simulated using the ns-3 network simulator, version 3.28 [33], with IEEE 802.11ax HEW indoor scenario model for dense WLANs (Typically suitable for office buildings). Some important simulation parameters are given in Table 5.1.

Table 5.1 MAC layer and PHY layer simulation parameters.

Parameters	Values
Frequency	5 GHz
Channel bandwidth	160/20 MHz
Data rate (MCS11)	1201/54 Mbps
Payload size	1472 bytes
Contention window minimum	32
Contention window maximum	1024
COSB design parameter (ω)	32
Simulation time	100/500 sec
Station position	Fixed/Random
Distance from AP	10/25 m
Propagation loss model	LogDistancePropagation
Mobility model	ConstantPositionMobility
Rate adaptation model	ConstantRateWifiManager MinstrelWifiManager
Error rate model	NistErrorRateModel YansErrorRateModel

To evaluate the QL parameter selection for the proposed *i*QRA, 25 contending STAs are simulated for 100 seconds, varying α and β with small (0.2), medium (0.5) and large (0.8) values. Probability ε was set to 0.5 for balanced exploration and exploitation. Figure 5.1 shows the convergence of learning estimate ΔQ from equation (4.29) with respect to the learning rate (α). The figure depicts how a smaller α makes ΔQ converge faster. The convergence of ΔQ indicates that the STA has learned its environment and can exploit optimal actions in the future. An interesting observation is that ΔQ is not steady in the beginning, which is due to the initial exploration of the environment. Therefore, most of the states do not optimize the Q-value function in the beginning. Later, the STA locates the states

that can deliver the most rewards, increasing the cumulative reward. After enough instances (13 instances for $\alpha = 0.2$ in Figure 5.1), it can be seen that the learner has found configurations that can lead to optimization of the process. Similarly, we observe in Figure 5.2 that ΔQ converges faster with the large value of β, compared to the smaller values. In both cases (Figure 5.1and Figure 5.2), ε was set to 0.5, indicating equal opportunities for exploration and exploitation. The small value for α and the large value for β (along with equal probability ε) yield the best results for optimization in the system. The convergence of learning estimates ΔQ shows that an optimal solution for the environment exists.

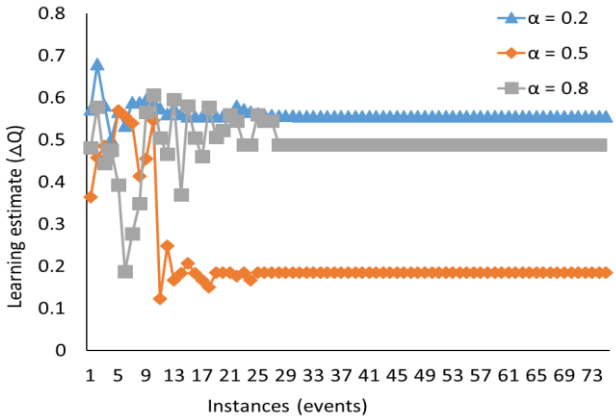

Figure 5-1 Convergence graphs of the learning estimate (ΔQ) from varying the learning rate α (β = 0.8).

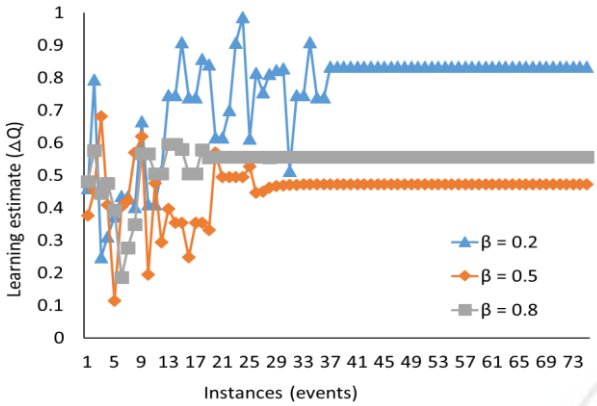

Figure 5-2 Convergence graphs of the learning estimate (ΔQ) from varying the discount factor β (α = 0.2).

Figures 5.3 to 5.8 portray the effects of the parameters on throughput of the system (Figures 5.3 to 5.5 for a small network of 15 STAs, and Figures 5.6 to 5.8 for a dense network of 50 STAs). As shown in Figure 5.3, if ε is set to 0.2 for a small network of 15 STAs, $\alpha = 0.5$ and $\beta = 0.2$ give the best results. However, in this case, decreasing α (that is $\alpha = 0.2$) has little effect on throughput, but increasing it to $\alpha = 0.8$ degrade throughput dramatically. Figure 5.4 shows that if ε and α are set to 0.5, β can be set small, medium, or large. However, for $\varepsilon = 0.8$ and $\alpha = 0.5$, seting β to its medium value (that is, $\beta = 0.5$) enhances throughput, as shown in Figure 5.5. Figures 5.6 to 5.7 show that for a dense network system of 50 STAs, a small value for α (that is, $\alpha = 0.2$) and a large value of β (that is, $\beta = 0.8$) are efficient for small and medium values of ε (that is $\varepsilon = 0.2$ and $\varepsilon = 0.5$). With a large value for ε (that is, $\varepsilon = 0.8$), as shown in Figure 5.8, throughput is improved if the large α and β are used (that is, $\alpha = 0.8$ and $\beta = 0.8$). Thus, from Figures 5.3 to 5.8, we show that a combination of a smaller α, a larger β, and a medium value for ε (that is, $\alpha = 0.2$, $\beta = 0.8$, and $\varepsilon = 0.5$) is somewhat efficient for both sparse and dense network systems.

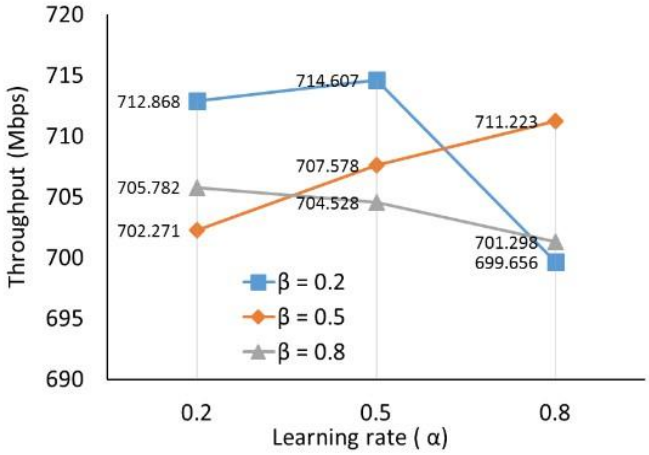

Figure 5-3 Throughput comparison of α and β in a small network of 15 STAs with ε = 0.2.

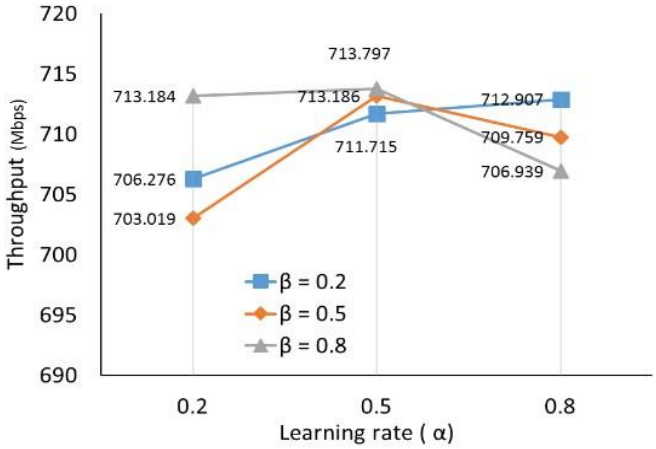

Figure 5-4 Throughput comparison of α and β in a small network of 15 STAs with ε = 0.5.

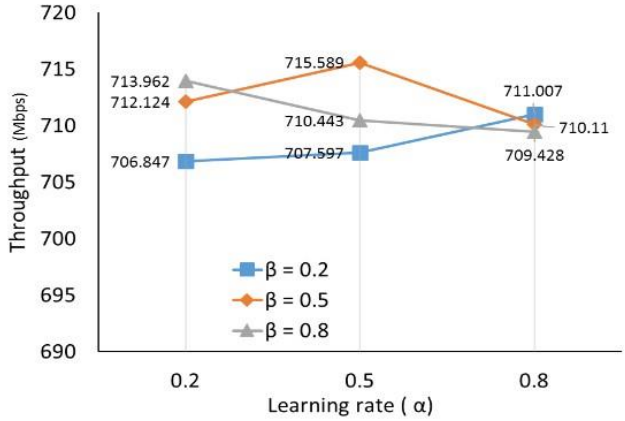

Figure 5-5 Throughput comparison of α and β in a small network of 15 STAs with ε = 0.8.

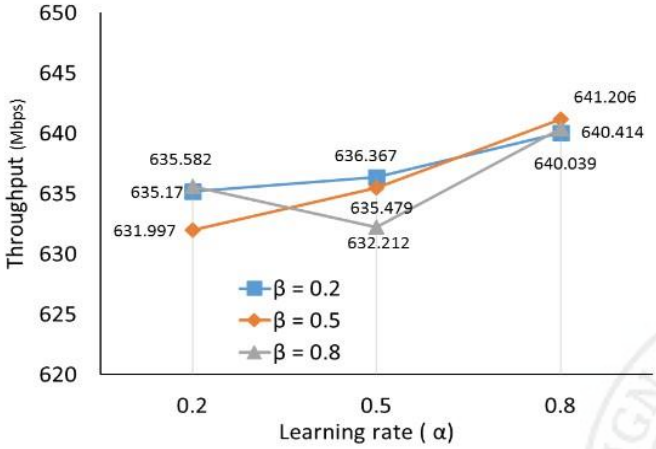

Figure 5-6 Throughput comparison of α and β in a dense network of 50 STAs with ε = 0.2.

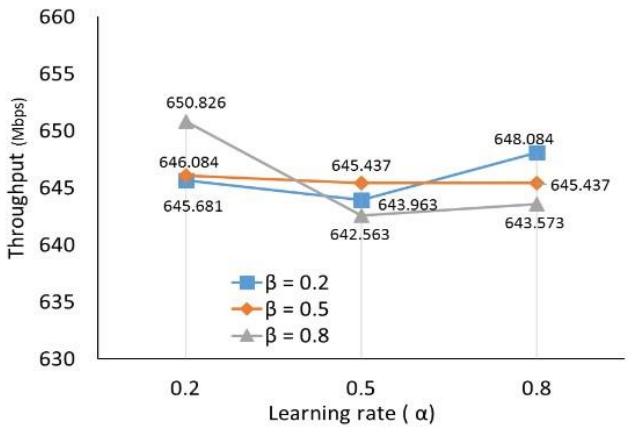

Figure 5-7 Throughput comparison of α and β in a dense network of 50 STAs with ε = 0.5.

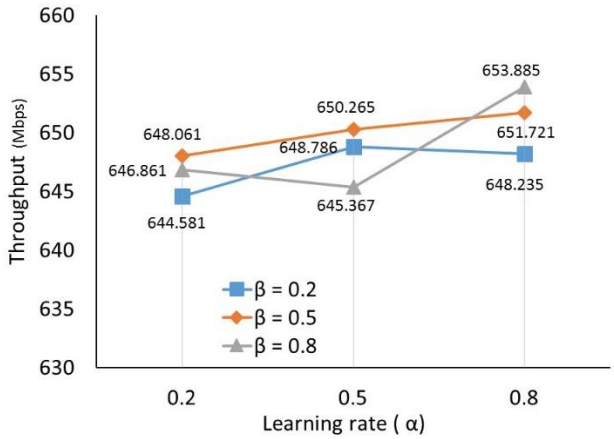

Figure 5-8 Throughput comparison of α and β in a dense network of 50 STAs with ε = 0.8.

5.2 Throughput

To evaluate the performance of *i*QRA, simulation results are compared with the state-of-the-art BEB and non-intelligent COSB algorithms. Figure 5.9 shows how the *i*QRA mechanism optimizes the throughput of COSB, specifically in a dense network of 50 contending STAs. The performance improvement clearly indicates that the QL-based proposed mechanism is effective at learning the wireless network. In a network of five contending STAs, *i*QRA achieves relatively lower system throughput than COSB. COSB outperforms for very small networks

(that is for less than 10 contending STAs). The performance of *i*QRA may degrades in small networks due to low and irregular rewards.

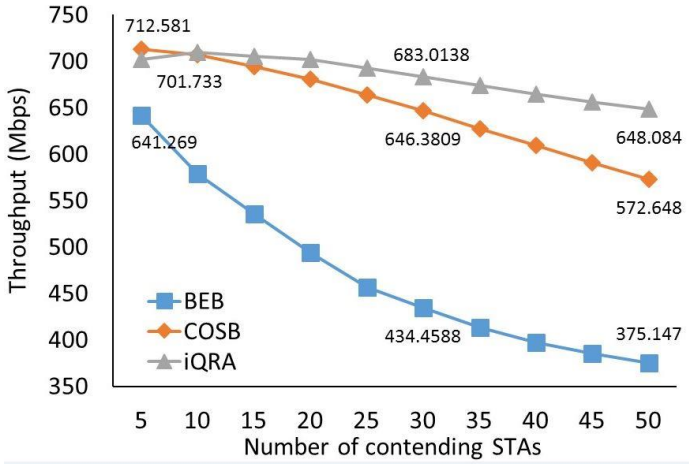

Figure 5-9 Throughput comparison of BEB, COSB, and *i*QRA with α = 0.2, β = 0.8 and ε = 0.5 in a network of five to 50 contending STAs.

5.3 Average Channel Access Delay

The channel access delay for a successfully transmitted data frame is defined as the interval from the time the frame is at the head of the queue (ready for transmission) until successful acknowledgement that the frame was received. If a frame reaches the given retry limit, it is dropped, and its time delay is not included in the calculation of channel access delay. Figure 5.10 depicts the performance of the proposed *i*QRA mechanism along with the conventional BEB and the original COSB mechanisms in terms of channel access delay (in milliseconds). From the figure, it can be observed that the proposed *i*QRA mechanism has a higher channel access delay, compared to COSB; however, it does not exceed the conventional BEB mechanism. It is obvious that the *i*QRA mechanism has an increased channel access delay due to its environment inference characteristics.

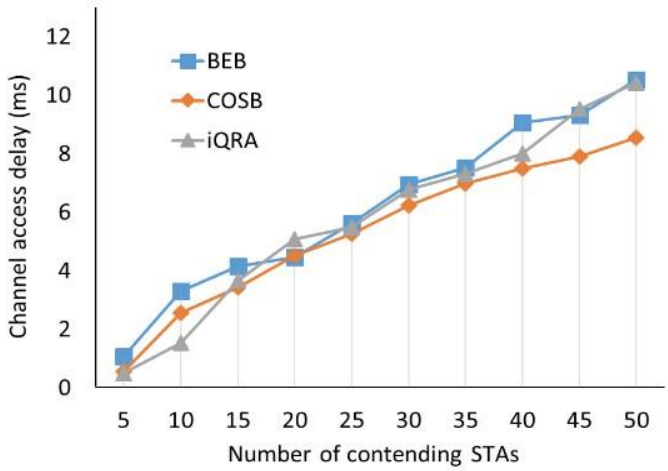

Figure 5-10 Channel access delay comparison of BEB, COSB, and *i*QRA with α = 0.2, β = 0.8, and ε = 0.5 in a network of five to 50 contending STAs.

5.4 Fairness

The fairness issue can be seen for COSB in Figure 5.11. In a dense network environment of 50 STAs, COSB suffers from the fairness problem due to some STAs continuously operating at a higher CW size, and a few fortunate STAs can operate at a lower CW size. Under COSB, once the STA reaches a larger CW, it has to transmit successfully many times to return to the smaller CW, which seems difficult in a dense network environment due to the high probability of collision. The proposed *i*QRA brings fairness to the contending STAs, because every STA autonomously and intelligently exploits its environment. Table 5.2 shows the values in Jain's fairness index [34] achieved by BEB, COSB, and *i*QRA for a small network of five STAs to a large dense network of 50 STAs. It is observed that the COSB mechanism performs unfairly for both small and large network environments, while the *i*QRA mechanism optimizes COSB to perform fairly among the contending STAs, whether it is for a small network or a large network.

Table 5.2 Jain's fairness index (JFI) comparison.

STAs	BEB	COSB	*i*QRA
5	0.999	0.953	0.999
10	0.999	0.999	0.999
15	0.999	0.999	0.999
20	0.999	0.997	0.999
25	0.999	0.992	0.999
30	0.998	0.993	0.999
35	0.998	0.991	0.999
40	0.997	0.990	0.999
45	0.998	0.948	0.999
50	0.998	0.953	0.998

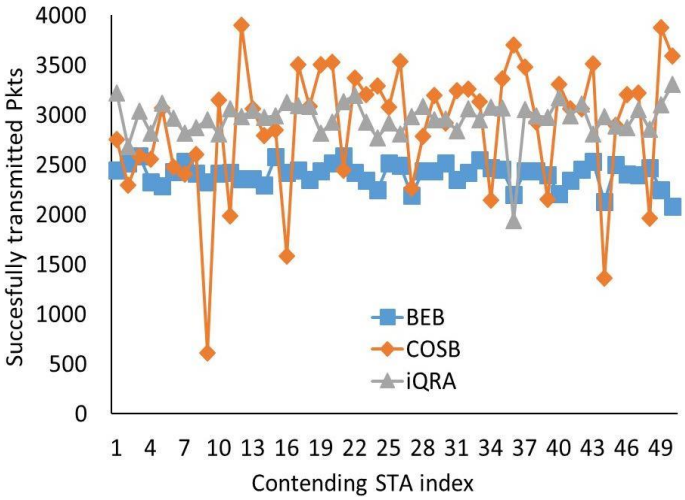

Figure 5-11 The number of successfully transmitted packets by each STA in a dense network of 50 STAs.

5.5 Network Dynamicity

Subsequently, QL is essentially intended to make intelligent adjustments according to the dynamics of the environment. A dynamic environment can be the activation of more contending STAs in the network or the deactivation of previously active STAs. This thesis evaluates the performance of the proposed *i*QRA mechanism by activating five more contending STAs every 50 seconds until the number of STAs reached 50. Figure 5.12 explains the effects of network dynamics on ΔQ (that is, learning estimate) of a tagged STA. The figure shows 1400 learning instances (events) of a tagged STA during the simulation period (500 sec). Each instance represents the updated value of learning estimate ΔQ

whenever a packet transmission is attempted. As shown in the figure, with changes in the number of contending STAs within the network, the tagged STA experiences a fluctuation in ΔQ, indicating the change in the environment. Later, this QL-equipped, intelligent tagged STA converges and is capable of optimizing the performance in a dynamic wireless environment. In Figure 5.13, it can be seen that *i*QRA eventually reaches a steady state in system throughput. On the other hand, BEB and COSB are severely affected by the increase in the number of competing STAs.

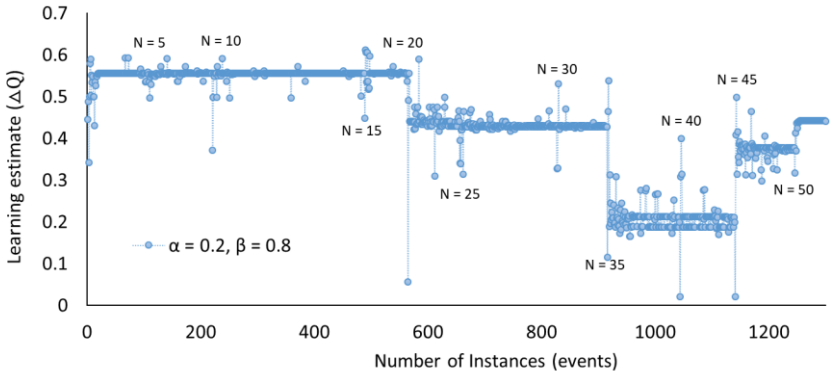

Figure 5-12 Convergence of the learning estimates (ΔQ) in a dynamic network environment (increasing the number of contenders every 50 seconds).

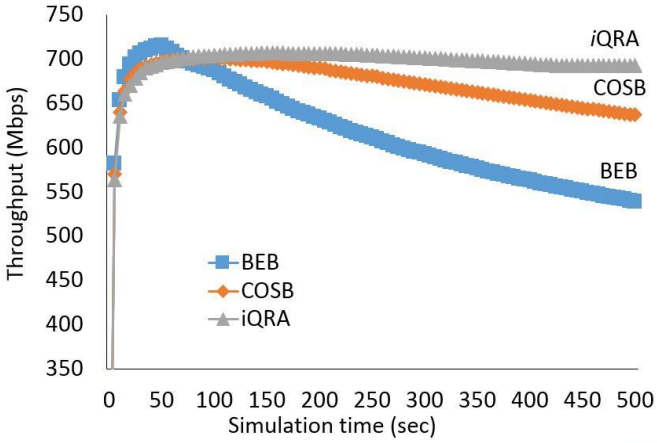

Figure 5-13 System throughput comparison in a dynamic network environment (increasing contenders by five every 50 seconds).

5.6 Distance-based Rate Adaptation Models

Throughput shown in Figure 5.9 and Figure 5.13 are achieved in a network environment using the ConstantRateWifiManager rate-adaptation algorithm [33], in which contending STAs are placed at a fixed distance from the access point (AP). Hence, all the devices are transmitting at a constant data rate. To evaluate the performance of the proposed *i*QRA algorithm, we simulated a more practical and real network environment, such as MinstrelWifiManager [33]. The Minstrel rate adaptation varies the transmission rate of the sender STA to match the WLAN channel conditions (mainly based on the distance from the AP), in order to achieve the best possible performance. The results shown in Figure 5.14 are achieved in an IEEE 802.11a (54 Mbps) wireless network for N = 10. All contending STAs were randomly placed within a distance of 25 *m* from the AP. A tagged STA (initially placed at a 1 *m* distance) moves away from the AP.

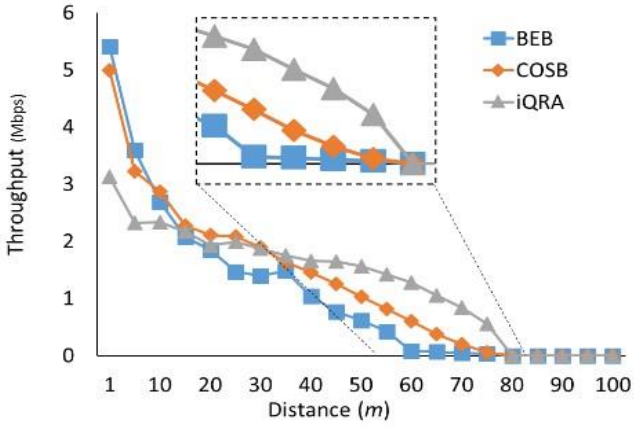

Figure 5-14 Throughput comparison for distance-based rate-adaptation network environment.

Throughput shown in Figure 5.14 was obtained after each 5 *m* distance from the AP. The performance of a tagged STA for all three of the compared algorithms (BEB, COSB, and *i*QRA) degrades as the distance from the AP increases, as shown in Figure 5.14. Observe that the throughput of the STA for BEB is close to zero after the STA reaches a distance of 60 *m*, and finally becomes zero when it exceeds the coverage (80 *m*). Under COSB, due to its observation-based nature, a STA

achieves higher throughput even after a 60 *m* distance, compared to BEB. However, the proposed *i*QRA maintains performance, even if the distance increases to 80 *m*, due to its intelligence capability.

5.7 Effects of Channel Error-Rate Models

In order to achieve reliable results to compare with real device performance, it is essential to represent the PHY layer of the WLAN as correctly as possible in simulations. The ns-3 simulator states two error-rate models for calculation of the bit error rate (BER) and corresponding packet error rate (PER): YansErrorRateModel and NistErrorRateModel [33]. Currently, ns-3 recommends using NistErrorRateModel as the default, specifically for ideal channel cases. There is not much difference between these two, except that YansErrorRateModel uses overly optimistic (analytical) results. In Figure 5.15, the effect of the above-stated error-rate models is evaluated. The figure shows that there is not much difference among BEB, COSB, and *i*QRA performance when simulated with the two different error-rate models. The performance of COSB increases a little with YansErrorRateModel. The reason is that, similar to YansErrorRateModel, COSB scales its parameters based on analytical results, that is, channel collision probability. On the other hand, the performance of *i*QRA remains almost the same, because it is the optimized form of COSB.

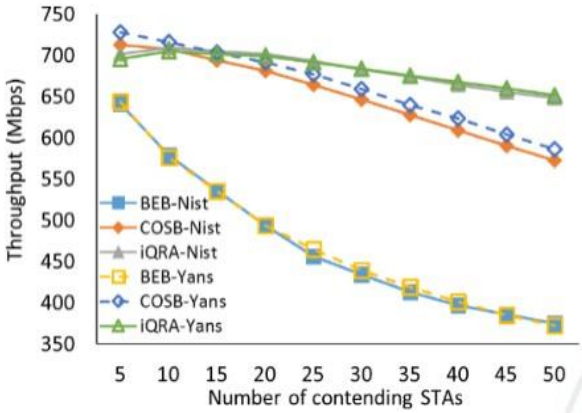

Figure 5-15 Throughput comparison for NistErrorRateModel and YansErrorRateModel network environments.

5.8 Multimedia data traffic (QoS-based traffic)

To reveal the strengths of proposed *i*QRA mechanism in QoS-enabled WLANs, we performed simulations with a QoS-supported IEEE 802.11 model for four multi-type of service data traffics with 54Mbps data transmission rate. Figure 5.16 shows the throughput comparison of the conventional BEB and the proposed COSB and *i*QRA for multi-type of service AC_BK and AC_BE access categories (AC). The figure clearly depicts that the performance of the ACs severely degrades with the increase of a number of contending STAs. Especially, the background data traffic type (BK) suffers much degradation due to less chance of channel access. Although the multimedia data types (that is VI, and VO) are of higher priority for channel access, their performance starts degrading as the number of contenders increases in the network as shown in Figure 5.17. The performance degradation with the increase of contenders depicts the blindness issue of currently implemented binary exponential channel access mechanism. As compared to the performance of BEB, the proposed *i*QRA outperforms for multi-type of service access categories, especially for BE, VI, and VO. However, the performance improvement is not much seen for BK data traffic type due to lowest priority data traffic in the network. The lowest priority of BK traffic allows the STAs to transmit less number of BK data frames, thus *i*QRA learns not much enough to optimize the performance of BK traffic. However, *i*QRA enhances the performance of BK data type in small size networks due to relatively less number of data frames from the other priority traffics as well. The proposed machine intelligence-based *i*QRA mechanism enhances the aggregate system throughput as shown in Figure 5.18. The performance improvement affirms the machine intelligence capabilities of the proposed mechanism.

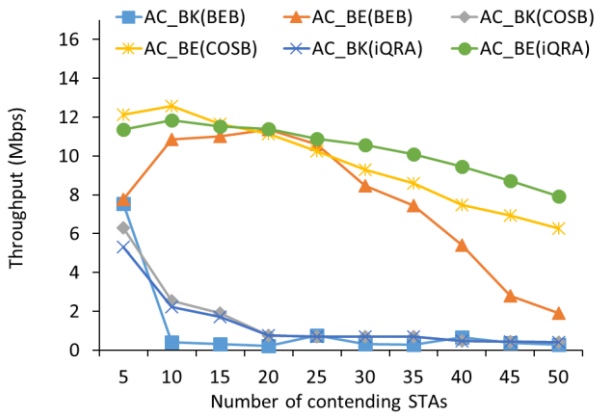

Figure 5-16 Throughput comparison of BEB, COSB and *i*QRA for AC_BK and AC_BE access categories.

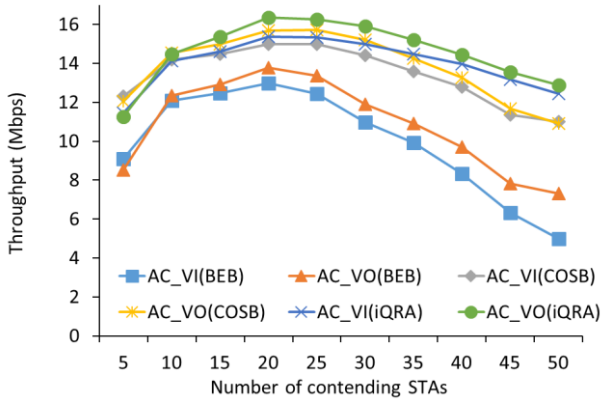

Figure 5-17 Throughput comparison of BEB, COSB and *i*QRA for AC_VI and AC_VO access categories.

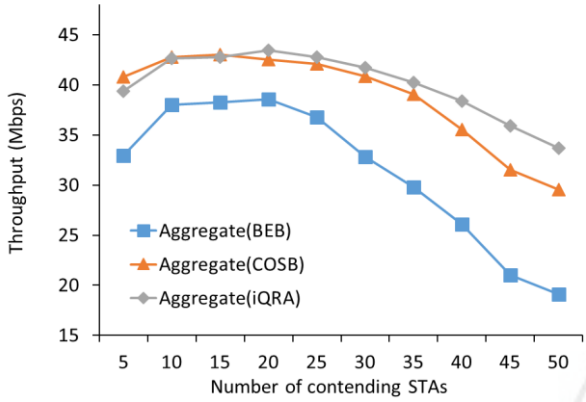

Figure 5-18 Aggregate system throughput comparison of BEB, COSB and *i*QRA in a QoS-enabled WLAN..

6. Conclusion

The upcoming dense high-efficiency WLAN (that is, IEEE 802.11ax HEW) promises four times higher throughput performance improvement of a single device. One of the bottlenecks for this performance achievement is tackling the huge challenge of efficient MAC layer resource allocation in WLANs due to their distributed contention-based nature. Currently, the CSMA/CA-based WLAN uses a binary exponential backoff mechanism (BEB), which blindly increases and decreases the contention window after collisions and successful transmissions, respectively. To handle the performance degradation challenge caused by the increasing density of WLANs, a self-scrutinized channel observation–based scaled backoff (COSB) mechanism based on a practical channel collision probability is proposed in this thesis. COSB overcomes the limitation of BEB to achieve high efficiency and robustness in highly dense networks, and enhances the performance of CSMA/CA in dense networks. However, to satisfy the diverse requirements of such dense WLANs, it is anticipated that prospective WLANs will autonomously access the best channel resources with the assistance of sophisticated wireless channel condition inference. Motivated by the potential applications and features of deep reinforcement learning (DRL) in wireless networks, such as the deployment of cognitive radio. In this thesis, one of the DRL techniques, Q-learning, is proposed as an intelligent paradigm for MAC layer resource allocation in dense WLANs. The proposed DRL paradigm uses intelligent QL-based inference to optimize the performance of COSB, named as *intelligent* QL–based resource allocation (*i*QRA). Simulation results show that the proposed *i*QRA mechanism optimizes the performance of COSB in fixed wireless STA network environments, as well as for randomly placed and distance-based rate adaptation network environments. It is also shown by the simulation results that proposed *i*QRA mechanism improves the performance of QoS-enabled WLANs as well, specifically for the dense network environments.

Future research considerations include the formulation of a mathematical model for the proposed *i*QRA mechanism. Future work also includes performance evaluations of *i*QRA in more realistic channel-error and signal-to-noise ratio (SINR)–based data rate models (that is diverse mobility models).

References

[1] IEEE 802.11ax High Efficiency WLAN (HEW), P802.11-TGax, 2014.

[2] R. Ali, S. W. Kim, B. Kim, and Y. Park, “Design of MAC layer resource allocation schemes for IEEE 802.11ax: Future directions,” *IETE Technical Review*, vol. 35(1), pp. 28-52, February 2018, DOI: 10.1080/02564602.2016.1242387.

[3] Ali, R., Shahin, N., Bajracharya, R., Kim, Y.T., Kim, B., and Kim, S.W. “A self-scrutinized backoff mechanism for IEEE 802.11ax in 5G unlicensed networks,” *Sustainability*, Vol. 10, no. 4, pp. 1201, 2018. DOI: https://doi.org/10.3390/su10041201.

[4] J. Moon and Y. Lim, “A Reinforcement Learning Approach to Access Management in Wireless Cellular Networks,” *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 6474768, 7 pages, 2017. DOI: https://doi.org/10.1155/2017/6474768.

[5] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen and L. Hanzo, "Machine Learning Paradigms for Next-Generation Wireless Networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98-105, April 2017. DOI: 10.1109/MWC.2016.1500356WC.

[6] Primer, Wi-Fi: Overview of the 802.11 Physical Layer and Transmitter Measurements. Beaverton: Tektronix Inc., 2013. pp. 4–7.

[7] M. X. Gong, B. Hart, and S. Mao, “Advanced Wireless LAN Technologies: IEEE 802.11ac and Beyond,” *GetMobile*, Vol. 18, no. 4, pp. 48–52, Jan. 2015.

[8] E. Charfi, and L. Chaariand L. Kamoun, “PHY/MAC enhancements and QoS mechanisms for very high throughput WLANs: A survey,” *IEEE Communications Survey and Tutorials*, Vol. 15, no. 4, pp. 1714–35, Nov. 2013.

[9] J. Govil, and B. Kumar, “Wireless LAN and IEEE standards,” *IETE Technical Review*, Vol. 23, no. 1, pp. 47–60, 2006.

[10] P. Barber. (2013). IEEE 802.11ax Project Plan, IEEE Technical Presentation [Online]. Available: http://www.ieee802.org/11/Reports/tgax_update.htm.

[11] K. Yunoki, and Y. Misawa, Possible Approaches for HEW, IEEE technical presentation, 2013 [Online]. Available: http://www.ieee802.org/11/Reports/tgax_update.htm.

[12] C. Zhu, Y. Kim, O. Aboul-magd, and C. Ngo, “Multi-User Support in Next Generation Wireless LAN,” in Proceedings of the 8th Annual IEEE Consumer Communication and Networking Conference, Las Vegas, NV, Jan. 9–12, 2011, pp. 1120–21.

[13] S. V. Bana, and P. Varaiya, “Space division multiple access (SDMA) for robust ad hoc vehicle communication networks,” in the proceedings of IEEE Intelligent Transportation Systems, 2001, Oakland, CA, Aug. 25–29, 2001, pp. 962–7.

[14] C. Guo, L. Zhao, H. Zhao, and W-E Chen, “A joint beamforming based SDMA protocol for IEEE 802.11n downlink,” in 2015 11th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QSHINE), Taipei, Aug. 19–20, 2015, pp. 166–70.

[15] IEEE MAC Enhancement for Quality of Service. IEEE Standard 802.11e, 2005.

[16] X. Yu, P. Navaratnam, and K. Moessner, “Resource reservation schemes for IEEE 802.11-based wireless networks: A survey,” *IEEE Communications Survey and Tutorials*, Vol. 15, no. 3, pp. 1042–61, Jul. 2013.

[17] M. A. Youssef, and R. E. Miller, “Analyzing the point coordination function of the IEEE 802.11 WLAN protocol using a systems of communicating machines specification, ” UMIACS Technical Report CS-TR-4357, UM Computer Science Dept., College Park, 2002, Vol. 36.

[18] IEEE Standard for Information Technology. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. ANSI/IEEE Std 802.11 2007, i-513.

[19] Er, M. J., Zhou, Y. “Theory and Novel Applications of Machine Learning,” InTech 2009, DOI: 10.5772/56681.

[20] Sutton R. S., and Barto A. G. Reinforcement Learning: An Introduction. Second ed.; MIT Press, Cambridge, MA, 1998, ISBN: 0262193981.

[21] Alpaydin E. Introduction to Machine Learning. Third ed.; MIT Press, Cambridge, MA, 2014, ISBN: 978-0-262-028189.

[22] Feng, V. S., and Chang, S. Y. “Determination of Wireless Networks Parameters through Parallel Hierarchical Support Vector Machines,” *IEEE Transaction on Parallel Distributed Systems*, Vol. 23, No. 3, pp. 505–12, 2012. DOI: 10.1109/TPDS.2011.156.

[23] Wen, C.-K., Jin, S., Wong, K.-K., Chen, J.-C., and Ting, P. “Channel Estimation for Massive MIMO Using Gaussian-Mixture Bayesian Learning,” *IEEE Transaction on Wireless Communications*, Vol. 14, No. 3, pp. 1356–68, 2015. DOI: 10.1109/TWC.2014.2365813.

[24] Yu, C.-K., Chen, K.-C., and Cheng, S.-M. “Cognitive Radio Network Tomography,” *IEEE Transaction on Vehicular Technology*, Vol. 59, No. 4, pp. 1980–97, 2010. DOI: 10.1109/TVT.2010.2044906.

[25] Xia M., Owada, Y., Inoue, M., and Harai, H. “Optical and Wireless Hybrid Access Networks: Design and Optimization,” *IEEE/OSA Journal of Optical Communications Networking*, Vol. 4, No. 10, pp. 749–59, 2012. DOI: 10.1364/JOCN.4.000749.

[26] Qiu, R. C., Hu, Z., Chen, Z. Guo, N., Ranganathan, R., Hou, S., and Zheng, G. “Cognitive Radio Network for the Smart Grid: Experimental System Architecture, Control Algorithms, Security, and Micro grid Testbed,” *IEEE Transactions on Smart Grid*, Vol. 2, No. 4, pp. 724–40, 2011. DOI: 10.1109/TSG.2011.2160101.

[27] Li, R., Zhao, Z., Zhou, X., Ding, G., Chen, Y., Wang, Z., and Zhang, H. “Intelligent 5G: When Cellular Networks Meet Artificial Intelligence,” *IEEE Wireless Communications*, Vol. 24, No. 5, pp. 175-183, 2017. DOI:10.1109/MWC.2017.1600304WC.

[28] Aprem, A., Murthy, C. R., and Mehta, N. B. “Transmit Power Control Policies for Energy Harvesting Sensors with Retransmissions,” *IEEE Journal of Selected Topics in Signal Processing*, Vol. 7, No. 5, pp. 895–906, 2013. DOI: 10.1109/JSTSP.2013.2258656.

[29] Alnwaimi, G., Vahid, S., and Moessner, K. “Dynamic Heterogeneous Learning Games for Opportunistic Access in LTE-Based Macro/Femtocell Deployments,” *IEEE Transactions on Wireless Communications*, Vol. 14, No. 4, pp. 2294–2308, 2015D. OI: 10.1109/TWC.2014.2384510.

[30] Ali, R., Shahin, N., Kim, Y.T., Kim, B., and Kim, S.W. “Channel Observation-based Scaled Backoff Mechanism for High Efficiency WLANs,” *Electronics Letters*, Vol. 54, No. 10, pp. 663-665, May 2018.

[31] Kosek-Szott, K., Gozdecki, J., Loziak, K., Natkaniec, M., Prasnal, L., Szott, S.,and Wagrowski, M. “Coexistence Issues in FutureWiFi Networks,” *IEEE Networks*, Vol. 31, No. 1, pp. 86–95, 2017.

[32] Ali, R., Zikria, Y.B., Amin, F., Kim, B., Kim, S.W. “I-DTMC: An Integrated-Discrete Time Markov Chain Model for Performance Analysis in Future WLANs,” In Proceedings of the 2017 IEEE 42nd Conference on Local Computer Networks Workshops, Singapore, 9–12 October 2017, pp. 160-164.

[33] The Network Simulator — ns-3. [Online]. Available: https://www.nsnam.org/.

[34] R. Jain, D. Chiu, and W. Hawe, “A quantitative measure of fairness and discrimination for resource allocation in shared computer system,” Eastern Research Laboratory, Digital Equipment Corporation, 1984.

Appendix A

SCI/SCIE Journal publications related to the dissertation

I. R. Ali, S. W. Kim, B. Kim, and Y. Park, "Design of MAC layer resource allocation schemes for IEEE 802.11ax: Future directions," *IETE Technical Review*, vol. 35(1), pp. 28-52, February 2018.

II. R. Ali, N. Shahin, R. Bajracharya, B. Kim and S. W. Kim, "A self-scrutinized backoff mechanism for IEEE 802.11ax in 5G unlicensed networks," *Sustainability*, vol. 10(4), pp. 1-15, April 2018.

III. R. Ali, N. Shahin, Y. Kim, B. Kim and S. W. Kim, "Channel observation-based scaled backoff mechanism for high-efficiency WLANs," *Electronics Letters*, vol. 54(10), pp. 663-665, May 2018.

IV. R. Ali, N. Shahin, Y. B. Zikria, B. Kim, and S. W. Kim, "Deep reinforcement learning paradigm for performance optimization of channel observation–based MAC protocols in dense WLANs," *IEEE Access*, (**accepted**).

SCI/SCIE/SCOPUS Journal publications not related to the dissertation

V. R. Bajracharya, R. Shrestha, **R. Ali**, A. Musaddiq, and S. W. Kim, "LWA in 5G: State-of-the-art architecture, opportunities, and research challenges," *IEEE Communications Magazine*, vol. 56(10), pp. 134-141, October 2018.

VI. M. A. Naeem, **R. Ali**, B. Kim, S. A. Nor, and S. Hassan, "A Periodic Caching Strategy solution for Smart City in Information-Centric Internet of Things," *Sustainability*, vol. 10(7), pp. 2576, July 2018.

VII. N. Shahin, **R. Ali,** and Y. Kim, "Hybrid Slotted-CSMA/CA-TDMA for Efficient Massive Registration of IoT Devices," IEEE Access, vol. 6, pp. 18366-18382, 2018.

VIII. Y. B. Zikria, **R. Ali**, R. Bajracharya, H. Yu, S. W. Kim,"IoT theoretical to practical: Contiki-OS and zolertia remote," *Far East Journal of Electronics and Communications*, vol. 17, no. 4, pp. 915-921, August 2017. (SCOPUS)

IX. **R. Ali**, S. Asghar, M. Spante, S. W. Kim, B.-S. Kim, "Competence Development at Workplace: A Work integrated informal Learning (WiiL) Approach," *Far East Journal of Electronics and Communications*, vol. 2, no. 2016, pp. 35-42, September 2016. (SCOPUS)

X. R. Bajracharya, R. Shrestha, **R. Ali**, and S. W. Kim, "Power saving Scheme for Wi-Fi Access Points," *Far East Journal of Electronics and Communications*, vol. 2, no. 2016, pp. 43-51, September 2016. (SCOPUS)

Conference publications not related to the dissertation

I. **R. Ali**, B. Kim, and S. W. Kim, "Machine-learning-based network-adaptable EDCA for QoS-supported wireless networks," *International Conference on Green and Human Information Technology*, Kuala Lumpur, Malaysia, Jan. 16-18 2019.

II. N. Shahin, **R. Ali**, and Y. Kim, "Performance Evaluation of Centralized and Distributed Control Methods for Efficient Registration of Massive IoT Devices," International Conference on Ubiquitous and Future Networks, Prague, Czech Republic, Jul. 3-6, 2018.

III. **R. Ali**, N. Shahin, A. Musaddiq, B. Kim, and S. W. Kim, "Fair and efficient channel observation-based listen-before-talk (CoLBT) for LAA-WiFi coexistence in unlicensed LTE," International Conference on Ubiquitous and Future Networks, Prague, Czech Republic, Jul. 3-6, 2018.

IV. M. Rafiq, R. Agyeman, H. K. Shin, **R. Ali**, K. M. Kim, G. S. Choi, "Deep Video Events (DVE): A Deep Learning Approach for Sports Video Summarization," The 13th IEMEK Symposium on Embedded Technology, Jeju, Korea, May 24-25, 2018.

V. N. Shahin, **R. Ali**, S. W. Kim and Y. Kim, "Adaptively Scaled Back-off(ASB) mechanism for Enhanced Performance of CSMA/CA in the IEEE 802.11ax High-Efficiency WLAN," IEEE/IFIP Network Operations and Management Symposium, Taipei, Taiwan, Apr. 23-27, 2018.

VI. **R. Ali**, Y. B. Zikria, F. Amin, B. Kim and S. W. Kim, "I-DTMC: An Integrated-Discrete Time Markov Chain Model for Performance Analysis in Future WLANs," 2017 IEEE 42nd Conference on Local Computer Networks Workshops (LCN Workshops), Singapore, 2017, pp. 160-164.

VII. A. Musaddiq, Y. B. Zikria, **R. Ali**, I. U. Rasool and S. W. Kim, "Congestion Control Routing Using Optimal Channel Assignment Mechanism in Wireless Mesh Network," in 2017 IEEE Ninth International Conference on Ubiquitous and Future Networks, Milan, Italy, July. 4-7, 2017.

VIII. **R. Ali**, M. U. Ali, W. Khalid, B. Kim, S. W. Kim, "Future WLANs: A Candidate for Smart Homes in Internet of Things (IoT)," In Proc. of International Conference on Advanced Computing, Communication and Information Science (ICACCI-2017), Daegu, Korea, May. 19-20, 2017.

IX. **R. Ali**, N. Hassan, B. Kim, and S. W. Kim, "The Diffusion of Internet of Things technology in Korea: A case study for technology adaptation analysis using Moor's Model," in the proceedings of International Conference on Information and Communication Technology and Digital Convergence Business (ICIDB-2016), Seoul, Korea, pp. 83-84, Dec. 16-17, 2016.

X. F. Amin, **R. Ali**, R. Bajracharya, and S. W. Kim, "Extended Advanced Zonal Rectangular LEACH (EAZR-LEACH) for Wireless Sensor Networks," in the proceedings of International Conference on Information and Communication Technology and Digital Convergence Business (ICIDB-2016), Seoul, Korea, pp. 61-62, Dec. 16-17, 2016.

XI. Y. B. Zikria, **R. Ali**, R. Bajracharya, H. Yu, and S. W. Kim, "IoT theoretical to practical: Contiki-OS and Zolertia Re-mote," International Conference on

Information and Communication Technology and Digital Convergence Business (ICIDB-2016), Seoul, Korea, pp. 179-184, Dec. 16-17, 2016.

XII. **R. Ali**, S. Asghar, M. Spante, S. W. Kim, B. Kim, "Social Networking sites for Interdisciplinary Learning at Workplace: Work integrated informal Learning (WiiL)," International Conference on Advanced Computing, Communication and Information Science (ICACCI-2016), Cebu, Philippine, May. 27-29, 2016.

XIII. M. I. Hussain, B. Ahmed, and **R. Ali**, "A discrete event simulation for the analytical modeling of M/D/1 queues: Output buffer of an ATM multiplexer," 2016 Sixth International Conference on Innovative Computing Technology (INTECH), Dublin, 2016, pp. 239-244.

XIV. R. Bajracharya, R. Shrestha, **R. Ali**, and S. W. Kim, "Dynamic Switch-Off Technique for Energy Efficient 802.11 Access Network," International Conference on Advanced Computing, Communication and Information Science (ICACCI-2016), Cebu, Philippine, May. 27-29, 2016.

XV. R. Bajracharya, R. Shrestha, Y. B. Zikria, **R. Ali**, I.U. Rasool, and S. W. Kim, "Green WLAN: Energy-Aware Access Point Management," International Conference on Information and Communication Technology and Digital Convergence Business (ICIDB-2015), Seoul, South Korea, Nov. 24 – 25, 2015.

XVI. **R. Ali**, S. W. Kim, B. Kim, "Performance of Network Redundancy in SCTP: Effects of Different Factors on Multi-homing," 2015 International Workshop on IT Convergence System (KJC-2015), Gyeongsan, Korea, Nov. 1-3, 2015.

고밀도 무선 네트워크의 매체접속제어 계층 자원 할당을 위한 딥러닝

알리라시드

영남대학교 대학원

정보통신공학과 정보통신공학전공

(지도교수 김성원)

요 약

인터넷 중심의 데이터 응용 분야에 무선랜(WLAN)이 광범위하게 사용되고 있다. 2020 년까지 전세계 인터넷 트래픽의 삼분의 이는 동영상이며, 트래픽의 절반 이상이 무선랜으로 전송될 것으로 예상되고 있다. 따라서 무선랜의 처리율과 효율 향상이 필요하다. 이러한 필요에 의해서 새로운 무선랜 기술들이 지속적으로 개발되고 있다. IEEE 802.11 표준화 위원회에서 제정된 IEEE 802.11ac 표준은 초당 기가비트의 전송률을 제공한다. IEEE 802.11ac 의 주요한 특징은 이전 표준보다 증가된 대역폭, 높은 변조율, 다중입출력(MIMO) 및 다중 사용자 전송모드이다. 무선랜의 효율을 증가시키는 기술이외에도 주파수 이용 효율을 높이기 위하여 IEEE 802.11ax 고효율 무선랜(HEW) 표준이 연구되고 있다.

사용자의 밀도가 매우 높은 차세대 환경에서 HEW 의 활용이 예상된다. 고밀도 환경에서 HEW 는 네배 높은 망 효율성을 보여준다. 하지만 현재의 무선랜은 시간 기반의

매체접속제어 자원할당(MAC-RA) 으로 인하여 효율적인 채널 사용이 제한되고 있다. 무선랜은 채널 접속을 위해서 이진 지수 백오프(BEB) 기반의 반송파 감지 충돌 회피 다중 접속(CSMA/CA) 방식을 사용하고 있다. BEB 에서는 충돌 구간(CW)를 위한 랜덤 백오프 값을 사용한다. CW 값은 전송이 실패하면 두배가 되고, 전송이 성공하면 최솟값으로 초기화 된다. 이러한 CW 값의 변화는 성능저하를 초래한다. 밀집된 환경에서 CW 값을 최솟값으로 초기화하면 충돌이 증가하여 망 성능이 저하된다. 작은 망 환경에서는 CW 값의 증가는 불필요한 대기시간을 증가시키게 된다. 밀집된 무선랜의 다양한 요구조건을 만족시키기고 충돌을 제어하기 위하여, 무선채널 환경에 따라서 자동으로 채널접속을 조절하는 방법이 필요하다. 미래 무선랜에서는 심화학습(DL)을 사용하여 이러한 지능적인 제어가 가능하게 된다.

IEEE 802.11 표준에 DL 을 적용한 기술은 탁월한 성능으로 인하여 미래 통신에 많이 적용될 것으로 예상되고 있다. 인지 무선과 통신 네트워크와 같은 다양한 분야에서 WLAN 의 MAC 계층에 DL 을 적용하는 연구가 진행되고 있다. 심화 강화학습(DRL)은 DL 의 한 기술이며, 주어진 환경에 반응하면서 학습자가 목적을 이루는 행동방식에 기반하고 있다. 본 논문에서는 밀집된 WLAN 의 MAC-RA 를 위한 DRL 기반의 지능적인 기술을 제안한다.

DRL 의 한 기술인 Q-러닝(QL)을 이용하여, 밀집된 WLAN 의 MAC-RA 를 위한 지능적인 QL 기반의 자원할당(iQRA) 기술을 제안한다. iQRA 는 백오프 매개변수(백오프 단계 및 CW 값)를 망 상황에 맞게 동적이며 자동적으로 조절하기 위하여 채널 관찰기반의 충돌 확률을 사용한다. 네트워크 시뮬레이터 3((ns3)를 사용한 모의실험에서 제안된 DRL 기반의 iQRA 기술이 다양한 WLAN 환경을 학습하고, 기존의 BEB 기술에 비하여 성능을 최적화하는 것을 확인하였다. 제안된 iQRA 기술의 성능을 다양한 WLAN 환경에서 측정하였으며, 처리율, 채널 접속 지연시간, 공평성이 성능지표로 사용되었다.