学位論文

Online Machine Learning Algorithms to Optimize Performances of Complex Wireless Communication Systems
（オンライン型機械学習アルゴリズムによる複雑無線通信システムの性能最適化に関する研究）

2021 年 9 月

大島 浩嗣

Abstract

This paper consists of six sections: introduction of this paper is in Chapter 1, the issues of current and future wireless communication systems are pointed out in Chapter 2, then the introduction of machine learning to overcome these issues are discussed in Chapter 3. The proposed two schemes, supervised learning based modeling and optimization based decisioning, and simple reinforcement learning based decisioning are introduced in Chapter 4 and 5 respectively, where the evaluation and the verification of the proposed schemes through real device experiments and computer simulations are also shown in these sections. Following are a brief summary of Chapters in this paper.

In Chapter 1, the background and the aim of this research are shown. Advancement of wireless communication technologies has brought us enormous positive change all over the world. Yet, from the viewpoint of exploiting its capability, there are still some issues to maximize their performance, which raises two simple questions. One question is that how to build models of today's and future complex wireless systems. Another question is that how to decide optimal action by using the models of wireless communication systems.

In Chapter 2, issues of today's and future wireless communication systems are indicated, especially in terms of optimizing their performance. Classical mathematical formulation based optimization scheme cannot be applied any more for today's complex wireless communication systems, because the complexity of the systems prevent building mathematical model. It opens the

window of applications of machine learning technologies to optimize performances of wireless communication systems.

In Chapter 3, application of machine learning technologies and its issues are discussed, then the proposed schemes in this research are introduced. Classical optimization of performances of wireless communication systems is based on mathematical formulation, which cannot be applied to today's complex, time-varying system usage. A data-driven modeling, by machine learning, is an aid for this issue. Deep reinforcement learning, deep learning based modeling and reinforcement learning based decision for action, is a state-of-the-art scheme in recent research fields. It has recently applied in the field of wireless communication a lot. However, it does not mean that all issues of current and future wireless systems can be solved by it. There still exist future works to be pointed out. One point is to seek alternatives for modeling to realize continuous function based modeling to obtain better solutions for continuous systems. Another point is to seek the feasible yet effective scheme if the amount of available information is small, like in IoT systems. Corresponding to these points, two novel schemes are proposed which are different not only from classical mathematical optimization but also from current state-of-the-art deep reinforcement learning approach.

In Chapter 4, one of the proposed schemes, supervised learning based modeling and optimization, is formulated and examined through experiments. It uses some amount of information to build the model of the wireless communication system, and obtain optimal parameters by an optimization algorithm. This is based on cognitive cycle using machine learning. It uses by supervised machine learning algorithm to build the performance model of the systems, obtains the optimal parameters by solving the optimization problem, takes action according to the decision, and updates the performance model in online manner. Two applications are shown: IEEE 802.11 WLAN and space communication. Through both real-world experiments and computer simulations, the validity of the proposed scheme is confirmed.

In Chapter 5, another proposed scheme, simple yet easily-implementable reinforcement learning, by MAB problem formulation is formulated and examined through experiments. By using a novel, light-weight, and distributed TOW algorithm, it realizes adaptive learning wireless communication systems whose capabilities in software and/or hardware are limited like IoT. Two applications are shown: heterogeneous network selection and channel selection in massive IoT. Through both real-world experiments and computer simulations, the validity of the proposed scheme is confirmed. These results show the effectiveness and feasibility of the proposed schemes.

This paper provides two novel approaches from wide viewpoint of current application of machine learning to wireless communication. The proposed scheme using supervised learning and optimization gives a better alternative of deep reinforcement learning especially when parameters are continuous. Another proposed scheme using simple reinforcement learning based on TOW, a ligh-weight MAB algorithm, provides feasible solution to increase performances of wireless communication systems where amount of available information is small like IoT. This research opened a new field of application of online machine learning technologies to optimize today's and future complex wireless communication systems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Recent advancement of wireless communication technologies brings the technologies to widely deployed and used all over the world. Along with widespread of usage, the requirements and use cases are becoming higher and broader, which leads to highly complex wireless communication systems.

Today's wireless communication systems are becoming large scale networks, where a large number of wireless devices are deployed in rather small area like massive IoT, and where the structures of networks are becoming more heterogeneous: multiple radio access technologies are used simultaneously in a mobile terminal, multiple radio transmission range are deployed, various requirements of application traffic are needed, etc. The complexity of wireless communication systems come from another way: the interaction among not only the functions inside wireless devices but also wireless devices through radio interferences. Moreover, the number of parameters to control in a device is large enough, due to the multiple layered structure of radio systems, which prevents the operators or systems to choose optimal values.

For example, for the last factor, there is the layered structure of wireless devices : physical layer handles the wireless link through transmission power control or , MAC layer deals with the management of wireless link by choosing appropriate modulation and coding scheme etc., also manages the access to wireless resource in time domain duplex (TDD) or carrier-sensing multiple access with collision avoidance (CSMA/CA). Network layer gives routing management and labeling of network address, typically internet protocol (IP). Transport layer handles the transportation of data on communication networks through the 2 way protocol like transport control protocol (TCP) or 1 way protocol like user datagram protocol (UDP). Session, presentation, and application layer has more according to user demand of application of communication networks.

These structures, namely open systems interconnection model (OSI model) of seven abstraction layers, are aimed to design as flexible as much in order to manage and meet various communication needs and demands with scalability. On those structures, the function among different layers are designed to be more independent, less-combined way for maintaining the flexibility and scalability. It brings flexible wireless communication systems such as IEEE 802.11 wireless local area network (WLAN) or 3GPP long term evolution (LTE) or 5G which can use common TCP/IP protocols on the internet.

On the other hand, just because the functions in each layer are designed independently, it is important but difficult to optimize end-to-end communication performance as a whole system. It can be easily understood just if considering the amount of wireless parameters is so large: modulation scheme of radio signals, selection of radio frequency and bandwidth, transmission power, and access scheme to wireless resources, etc. In addition, as mentioned above, higher layer protocols should be included to optimize user experience. It means that the conventional optimization of the performance of wireless communication systems are not able to be applied any more, because it is based on mathematical formulation which is too complex to be realized. It is essential to solve this issue in order to optimize the performance of future wireless system like Beyond 5G/6G, which is and will be becoming more and more complex.

The aim of this paper is to find a novel engineering scheme to overcome this issue by introducing a new technologies. It introduces machine learning technologies, which are data-driven modeling approaches, and have been developed very recently to solve issues in various fields of the society. In the field of wireless communications, some applications of machine learning technologies have been researched recently, such as supervised learning for signal processing, deep learning to predict the traffic demand and emphasize the quality of experience (QoE), etc. These researches mainly focus in a certain layer or function, which seems to be not enough to optimize the end-to-end communication quality as a whole. It also can be said that they are tend to be theoretical analysis and the implementation to real wireless devices or the operation of them are not researched yet. Indeed, to the best of my knowledge, there are less than 10 % of IEEE published papers which include whole wireless communication system optimization and the verification of prototyping of them, shown in Fig.1.1.

The performance of wireless communication systems are to be defined as an end-to-end communication quality, which requires the optimized set of parameters of wireless devices from in physical layer protocol to application layer protocol, and also required the optimized network among wireless/wired communication nodes. Considering using machine learning technologies, the more date is preferable to build more precise model of wireless communication systems, and take better actions. However, in the real world equipments, it is not always realized such data-rich environment: wireless devices are fundamentally short in their resources such as limited amount of battery, limited amount of central processing unit (CPU) power, namely, computational resources, and limited amount of parameters to control due to software or hardware restrictions.

It can be said that it is crucially important to seek an implementable, highly flexible, and higher capable wireless communication systems and its basic scheme to optimize them by using machine learning technologies in today's and future wireless communication systems like beyond 5G or 6G.

1.2 Purpose of Research

Based on the understanding of current status and issues indicated previous section, this paper proposes an online machine learning algorithms to optimize performances of complex wireless communication systems. These enable wireless communication systems to adapt the environment and the demand of traffic autonomously and intelligently in online manner. Applications including implementation on the wireless devices and verification through both real world experiments and computer simulations are also shown.

It is expected that more data obtained and more parameters can be controlled, intelligent wireless communication systems using machine learning can take better actions and reach higher performance in various environments and use cases. However, there are lots of limitations in real world devices, in terms of quantity and quality, which come from the software and hardware capability. Deep learning technologies [1] are novel applications and very strong tools for data-driven modeling for various systems and various purposes. It is an application of supervised learning and recently became famous of image recognition and Google AlphaGo [6]; in the latter case the combination of deep learning and Q-learning, deep reinforcement learning (DRL) was used. While many literatures has recently applied DRL for the wireless communication field [150–288], there are still challenges and open issues: How to implement on and operate with real devices on condition that they process the large amount of data based on high computational capacity. It would be hard to realize enough if considering mobile devices which have rather short resources of hardware.

Considering those issues, this paper proposes a new schemes of optimizing the performance of complex wireless communication systems using machine learning technologies. The main proposal is composed of two schemes based on the mapping of conventional and existing optimization schemes: one scheme is a supervised learning based modeling and optimization, another scheme is a simple reinforcement learning based optimal decision using a multi-armed bandit (MAB) algorithm.

The first scheme is based on supervised modeling like DRL, but the algorithm uses rather continuous function, and the optimization algorithm is introduced which is similar to the conventional theoretical and mathematical optimization approaches, not like in the trial-and-error based reinforcement learning approaches. It would be more suitable if the relation among parameters and resulting performances is rather continuous, such as network throughput performance and parameters like transmission power or the length of waiting time etc. In this paper, as an example of continuous function, support vector regression (SVR) and particle swarm optimization (PSO) algorithm are introduced for the decision of optimal parameters in feasible time. In order to validate the possibilities of real world implementation, these algorithms are implemented on IEEE 802.11 wireless local area network (WLAN) devices and experiments using these devices are examined. The computer simulations are also conducted to verify the scalability of proposed scheme. In addition, a different application for space communication area, emulator-based network experiments are conducted and verified.

The second scheme is a simple reinforcement learning approach, which is based on the formulation of MAB problem and uses a novel MAB algorithm called tug-of-war (TOW). It can be applied to light-weight devices like massive IoT. Those devices require distributed algorithms to optimize their decision, especially in time-varying situation. Simple reinforcement learning, based on the MAB problem formulation, can satisfy the requirements. In this paper, as examples of applications of this scheme, heterogeneous network selection is examined with real wireless devices. Also, Zigbee network in dense use case is examined through computer simulations.

1.3 Outline

The rest of this paper are following: starting with the indication of issues of current and future wireless communication systems in Chapter 2, various approaches of the optimal decisioning by machine learning based modeling are reviewed in Chapter 3 and the proposed schemes are indicated. In Chapter 4, one of the proposed schemes, supervised learning modeling and optimization algorithm, is elaborated and some experimental results are shown. In Chapter 5, another scheme, simple reinforcement learning based optimal decision using an MAB algorithm called tug-of-war (TOW), is elaborated and some experimental results are shown. Both chapters include implementation on real world devices and experiments. Finally, in Chapter 6, conclusion and some remarks are described.

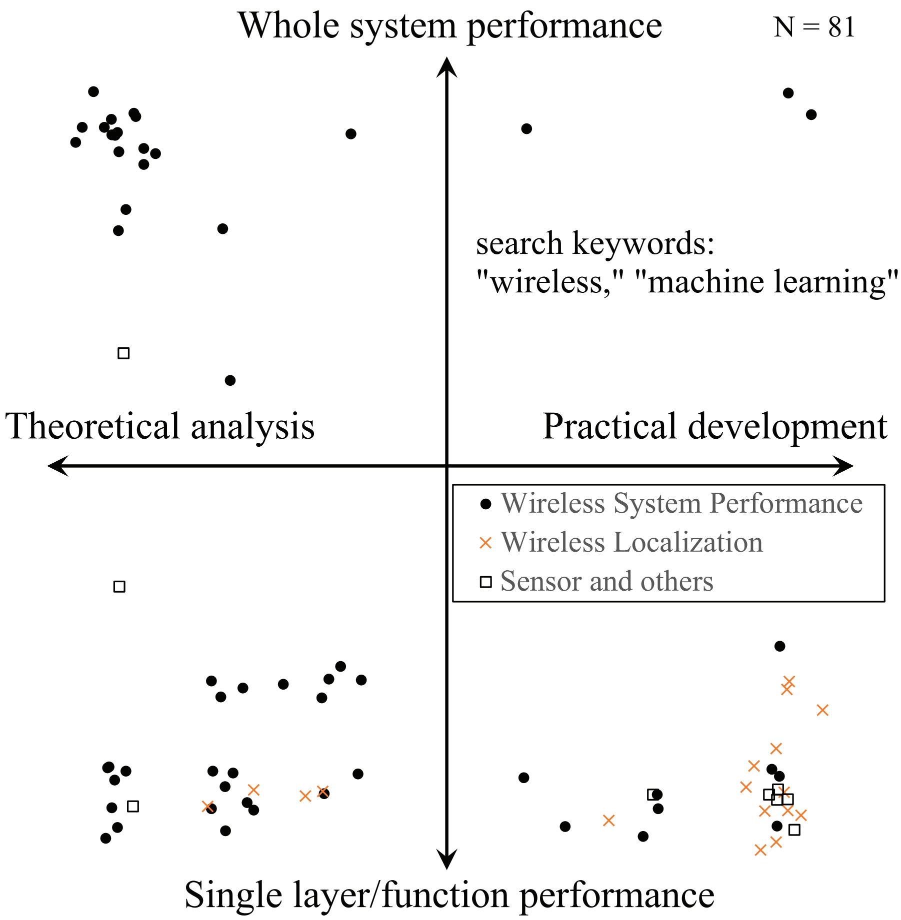


Figure 1.1: Categorization of search results from IEEE Xplore journal papers (cited at least once from others) with keywords “wireless” and “machine learning”.

Chapter 2

Issue of Current and Future Wireless System

2.1 Introduction

5G system [18] has just been deployed several regions in the world very recently. Comparing to 4G/LTE or 3G, the major difference of 5G system is that there is no single technology to develop like OFDM in 4G/LTE or CDMA in 3G.

A 5G system network consists of radio access network (5G NR) and core network (5G CN). For 5G NR, physical transmission technologies are proposed such as OFDM(A) on higher frequency band like 60 GHz millimeter wave, massive MIMO with beam foaming, and so on. For 5G CN, several network management technologies are required such as network function virtualization (NVF), software defined network (SDN), edge computing, and network slicing. Along with them, 5G system is more comprehensive: it can include legacy 4G/LTE system, licensed assisted access (LAA) / licensed shared access (LSA) assuming the usage of industrial, scientific, medical (ISM) band frequency which requires a certain channel access scheme for sharing spectrum with other wireless systems like listen before talk (LBT).

From the view point of performance requirements, there are three major requirements for 5G system: enhanced mobile broad band (eMBB), massive machine type communications (mMTC), and ultra-reliable low latency communications (URRC). Fundamentally, these are different aspects of requirements each other. Indeed, as use cases, various independent scenarios are proposed based on these requirements: a rapid download of large data such as movie data of some gigabytes (GB) by satisfying eMBB, a management of massive amounts of mobile sensors like IoT devices by satisfying mMTC, and telemedicine with extreme low latency wireless network by satisfying URRC.

To satisfy these extreme requirements of 5G system, high performance devices/equipments are needed in the wireless communication systems, while being capable of managing low-end devices like IoT. All these requirements and use cases means, the current and future wireless communication systems are becoming more and more complex and heterogeneous.

2.2 Research challenge to optimize future wireless communication system

2.2.1 New era of machine learning equipped wireless communication systems

Several literatures indicate that the management of 5G/Beyond 5G system requires machine learning technologies [26, 51–53, 55, 64, 149]. For example, in [149], applications of machine learning in wireless communication systems are shown as in Table 2.1.

2.2.2 Cognitive radio technology

Cognitive radio [3] is a classical but fundamental concept of a wireless system which can adapt the various change of environment: sudden increase or decrease of traffic demand, variation of wireless channel, contention among wireless transmitters, and so on. The key idea of cognitive radio is learning

Table 2.1: Machine learning algorithm to enhance cellular networks (modified from [149]).

Function	Examples	Algorithms
Sensing	Detection of network anomalies or events by multiple-entry data from hybrid sources	Logistic Regression (LR), Support Vector Machine (SVM), Hidden Markov Model (HMM)
Mining	Classifying services according to the required provisioning mechanisms (e.g., bandwidth, error rate, latency)	Supervised learning: - Gradient Boosting Decision Tree (GBDT), Unsupervised learning: - Spectral Clustering, - One-class SVM, - Replicator Neural Networks (RNN)
Prediction	Forecasting the trend of UE mobility or the traffic volume of different services	Kalman Filtering (KL), Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), Deep Learning (DL): - Recurrent Neural Networks (RNN), - Long-Short Term Memory (LSTM), Compress Sensing (CS)
Reasoning	Configuration of a series of parameters to better adapt services	Dynamic Programming (DP): - Branch-and-Bound Method, - Primal-and-Dual Method, Reinforcement Learning (RL): - Actor-critic Method, - Q-Learning Method, Transfer Learning (TL)

and action, and the feedback cycle of them. This concept is very similar to that of machine learning, especially online machine learning like reinforcement learning. It observes environmental variables, peformance, then makes a policy to adapt the change in environment, then makes action according updated policy, then observes the result to update the policy more sophisticated one.

It is clear that this concept, cognitive radio or cognitive cycle [4], will be suitable to the complex future wireless communication systems. In addition, recent advancement in the field of machine learning technologies can boost the progress of cognitive radio development through realizing and implementing the key idea, learning, to the wireless systems in real world.

In this paper, according to the above mentioned viewpoints, the application of machine learning technology, especially online manner, to optimize complex wireless communication systems are researched.

2.3 Summary

In summary, current 5G or beyond 5G/6G communication systems bring about the difficulties, to optimize their performance, which are consist of various protocols and layers inside a communication node, heterogeneity of a communication network, and the interaction among them and environment, namely, their complexity as a whole end-to-end system. As a result, some new technologies are required to improve the whole system performance. In this paper, the applications of machine learning technology is proposed and discussed through categorization of application to optimize complex wireless communication systems. Next section elaborates that point.

Chapter 3

Machine Learning for Wireless System

In this section, various machine learning application for wireless communication systems are reviewed in terms of the amount of information available in each system model, the type of decision of best action within each system. Through comparing to conventional and current research works, the approach of this paper is indicated clearly.

To begin with the conclusion of this section, there are two major issues in current and future complex wireless communication systems. These arise from the simple question: How the wireless nodes decide optimal action in today's complex wireless communication systems. The issues are following:

- Issue 1. Classical one-way optimal decision is not suitable for complex, time-varying situations any more. We have to seek the alternatives.

- Issue 2. Classical mathematical formulation and optimization technique can not be applied to complex, time-varying situations any more. We have to try other approaches.

3.1 Cognitive radio

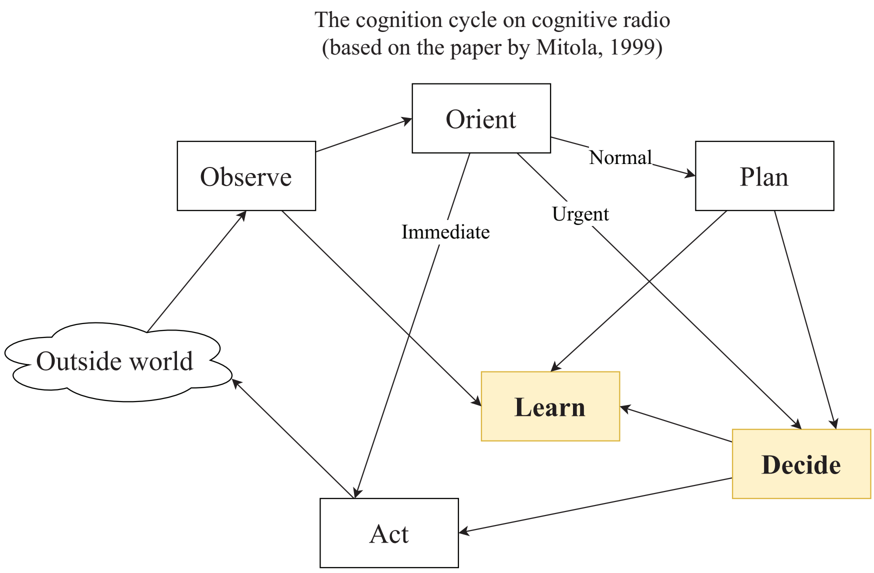
Cognitive radio technology, proposed by Mitola in [3] in 1999, is a fundamental concept of intelligent wireless communication system. Fig.3.1a shows the cognition cycle proposed by Mitola [3]. It learns environment and the behavior of wireless communication nodes, make trial-and-error to seek best action to improve the performance. Haykin showed a cognitive cycle of intelligent radio in [4]. Fig.3.1b shows the cognitive cycle proposed by Haykin [4]. An intelligent radio shown in this figure observes wireless environment, conducts radio-scene analysis, and estimates channel state and builds the predictive model. It uses this model to estimate channel capacity and control the transmission power and manages the spectrum to use. It then transmits radio signals, which in consequence becomes feedback for the radio itself.

These works did not refer to examples of learning algorithms though, still it is a comprehensive model of adaptive and intelligence modern wireless communication system.

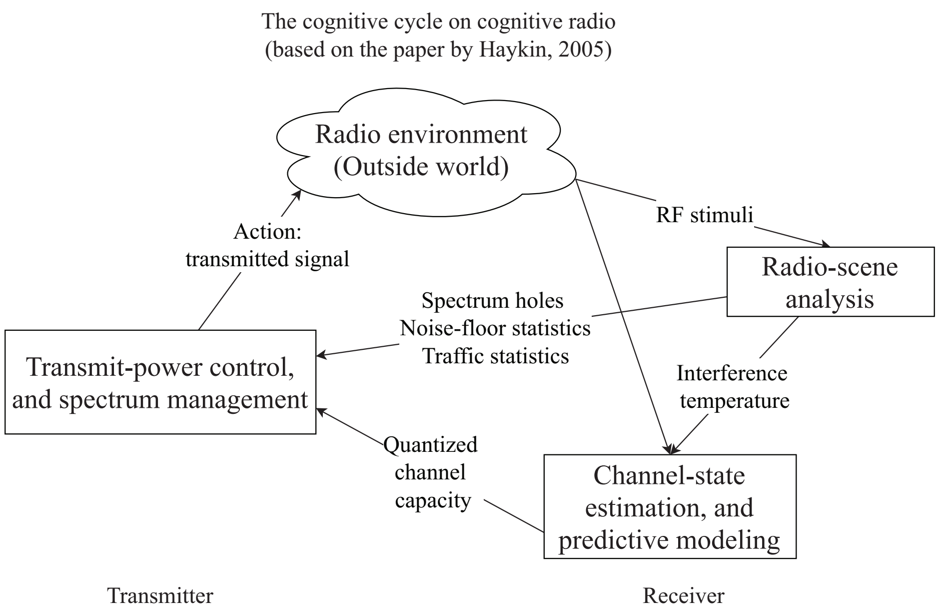
3.2 Classical Optimization and Current Machine Learning Approach

There has been a number of literatures seeking the optimal decision of wireless communication systems through mathematical and theoretic formulation of wireless channel and transmission power control [39]- [44], modulation and coding, the behavior of MAC protocol [2] or higher layer protocol, etc. The common approach of these researches is to define the mathematical model of the function of wireless communication system, then to formulate the maximization (or minimization) problem, and to obtain the optimal solution by solving the problem.

A good point of these "classical" approach is that the theoretical optimal solutions or parameters, or at least upper- or lower-bound of them can be obtained, under the assumption of the continuity of the function described.



(a) Concept of cognitive radio [3]



(b) Cognitive cycle of cognitive radio [4]

Figure 3.1: Concept of ognitive radio and cognitive cycle.

On the other hand, in terms of the modeling of wireless communication systems, the classical approach generally focused on a certain layer performance such as channel capacity in physical layer or throughput in MAC layer, and cannot cover whole system modeling. Indeed, current and future complex wireless communication systems are hard to be described mathematically as a whole system. Due to this limitation, the classical approach faces fundamental difficulty in applying the optimization of modern wireless communication systems.

Machine learning technologies, which are becoming more and more important solution for various issues in current society, could cope with this difficulty. They give the strong tool to build a whole system model by using numerous collection of data in a wireless communication system and wireless communication network. Machine learning technologies, includling deep learning and its relatives, have been more and more researched in the field of communication technology very recently [51–148].

3.3 Optimal Decision by using Machine Learning Technologies

When using machine learning to optimize the decision and action of wireless communication systems, there are two points of view. One point is the amount of data. Supervised learning, especially deep learning and its relatives need and can deal with numerous amount of data, to extract the characteristics of the system from which the data is collected. If the amount of data is limited, then reinforcement learning approach would be more suitable. It can make optimal decision through the iteration of trial and error cycle under rather the environment of limited information and parameters to control.

Another point is to how to decide the optimal action to achieve higher performance. There are two strategy: decision by learning scheme or by optimization algorithm, namely maximization or minimization of a formula. If the change of the environment surrounding wireless communication systems are relatively slow, and if the relation between parameters and performance is continuous, not discrete, then optimization algorithm would be suitable. It is also notable that if the relation between parameters and performance is discrete, then reinforcement learning approach would be suitable as well.

Fig.3.2 shows the relation of these approaches and examples of application of optimizing wireless communication systems. The approaches of the research in this paper are on two panes of this figure. One on the upper-left pane is a sort of simple reinforcement learning, using MAB, multi-armed bandit problem formulation. Another on the lower-right pane is well-known cognitive cycle based approach using machine learning (mainly supervised) and optimization algorithm. To the best of my knowledge, former researches by the latter approach, are hardly found. These are elaborated in the following sections.

3.3.1 Classical theoretical model and optimal decision

The lower-left pane in Fig.3.2 indicates "classical" theoretical model and optimization. It uses mathematical formulation of wireless communication systems, usually focusing on a certain layer (or two layers like physical layer and MAC layer). Then the problem is converted to the optimization problem: optimization are applied for maximization or minimization of the equation formulated above, which gives the optimal parameters of the equation. The obtained optimal parameters are, under the assumption of continuity of the function expressed in the formulated equation, theoretically optimal values by definition. There many good examples of this type of researches, such as water-filling optimization of transmission power [290].

3.3.2 Deep learning based model and reinforcement learning based decision

Deep learning (DL) is a newly developed and a rapidly spreading technique in various fields in the current society. It is an advanced form of a artificial neural network, and a kind of supervised learning. The first achievement of the DL was in the field of computer vision. It seems to simulate a human image recognition, and the impact of that spread all over the world.

In the filed of wireless communication systems, DL technique has been introduced in various layers of communication [55]. The early applications of DL are in the estimation of parameters of propagation channel [65, 85, 258], device location estimation [68, 101, 113, 113, 116, 117, 139], etc.

Deep reinforcement learning (DRL) has been applied very recently in the field of wireless communication systems. DRL is a combination of deep learning and reinforcement learning as shown in Fig.3.3. It can be said as an implementation of cognitive cycle: it learns the relation among environment, parameter, action, and performance of the wireless communication node by deep learning. It also decides its action by seeking better action through trial-and-error: reinforcement learning. The major strong point of DRL is to build the performance model by deep learning in online manner, and to utilize it to predict the performance of the systems when certain parameters are deployed. Reinforcement learning, usually Q-learning based algorithm, is applied to seek better actions by evaluating the results, update its network, and choose predicted better parameters by using deep learning. Note that DRL always requires the information of the state of wireless communication systems, which might unrealistic in the real world.

3.3.3 Supervised learning based model and optimal decision

DRL is a strong technique to seek better decision of wireless communication systems as described in the previous section. However, there are some assumptions and limitations. First, it needs a large amount of data due to the training in deep learning. It leads some temporal training overhead in the real world implementation. Second, it uses reinforcement learning: even if the variables are continuous and are able to be optimized by minimization or maximization of continuous functions, it is forced to do trial-and-error processes. It may suffer form insufficient performance due to the fact that the number of trial is finite in the real world. In other words, the DRL approach can be underperform than the classical mathematical formulation and optimization approach. It is because the classical scheme brings about theoretically determined parameters which will achieve maximum or at least some sort of upper-bound performance of wireless communication systems. It can not be assured in general that the DRL approach reaches the theoretical maximum performance.

These discussion give an insight of a better solution of using machine learning technologies to optimize wireless communication systems. What if some sort of mathematical optimization can be applied to seek the best parameters, while using machine learning as a tool to build the performance model of wireless communication systems? The answer proposed in this paper is the wireless communication systems optimization method based on cognitive cycle using machine learning and an optimization algorithm. It uses a supervised (or unsupervised, depending on the problem) learning to build the performance model of wireless communication systems, and defining the optimization problem using this performance model as a function of variables, observables, and performance of the systems. Then, by solving that optimization problem, the optimal parameters are obtained. After taking actions according to the optimum parameters, wireless communication systems observe the results, then update the performance model by the machine learning. This feedback loop is an implementation of a cognitive cycle based on machine leaning, taking advantages of classical mathematical formulation approach. The detail of the proposed scheme is elaborated in the following section. Note that the word "optimization" used here means mathematical optimization: the maximization or minimization of certain functions, not in a strict meaning of simple optimization problem.

3.3.4 Simple Reinforcement Learning Decision

The assumption of the proposed cognitive cycle approach described in the previous section is that the communication nodes in wireless communication systems to be optimized can obtain and control various observables and parameters. However, wireless equipments in the real world, especially simple devices like IoT sensors, can not deal with such multiple parameters in general. Indeed, one or a few parameters to control, the same as observables, are typical in IoT devices. In addition, the computational resources and communication methods are limited in such devices. So, more simple, light-weight algorithms are worth to be developed.

The multi-armed bandit (MAB) problem [19] is a simple machine learning problem, while achieving good performance in the limitation of finite number of trials. It is used in some area of wireless communication systems, like a channel selection in a cognitive radio [7, 8], or the resource allocation in 5G small cell [289], but the numbers of researches is limited. In addition, very a few or none of them show the experimental validation of the research by implementing on wireless devices.

In this paper, a light-weight and high performance algorithm is proposed using the Tug-of-War (TOW) algorithm which has been developed recently. It can achieve as high performance as well-known algorithm like UCB-1 tuned, while the implementation is very simple [23–25]. It is also worth to be noted that the TOW does not require any information of the current state of the system.

3.4 Summary

In summary, this paper proposes two novel approaches which are different from classical mathematical optimization and current state-of-the-art deep reinforcement learning. The first proposed approach is that the wireless communication systems optimization method based on cognitive cycle using machine learning. It uses machine learning to build the performance model of the systems, obtain the optimal parameters by the optimization formulation, and update the performance model in online manner. The second approach is a simple reinforcement learning, MAB problem formulation. By using a light-weight algorithm, it is more feasible in terms of the implementation and the operation of wireless devices like IoT.

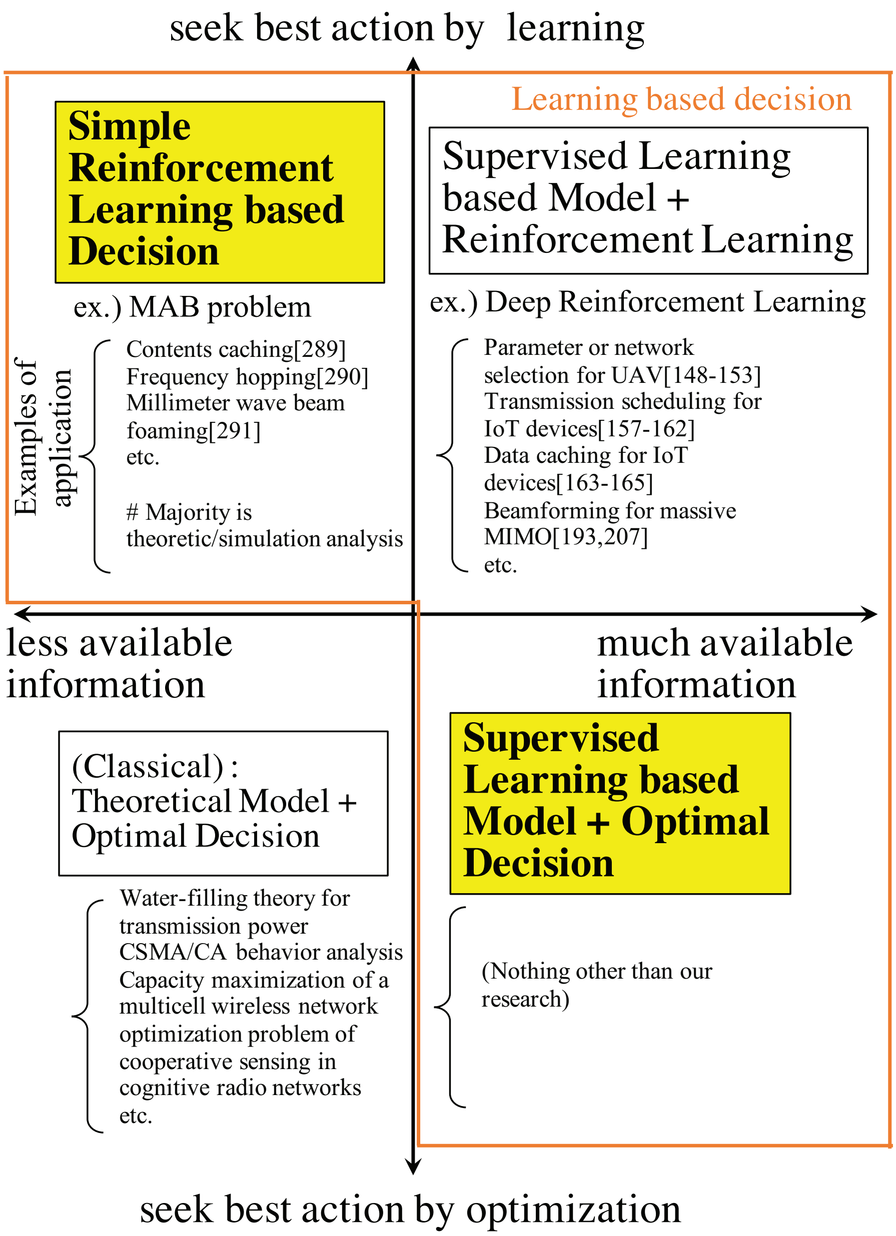


Figure 3.2: Optimization strategy of wireless communication systems using machine learning.

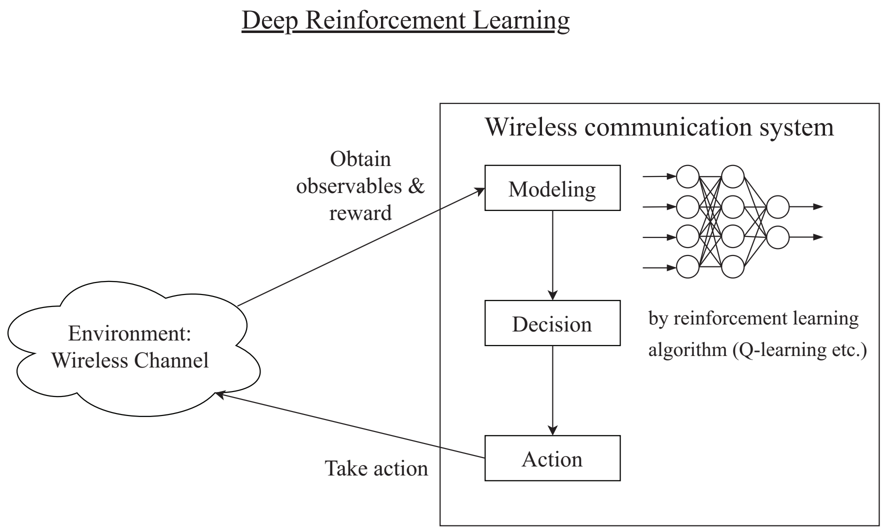


Figure 3.3: The concept of deep reinforcement learning.

Chapter 4

Optimal decision through cognitive cycle using supervised learning based model

Wireless traffic and the number of wireless communication devices have increased rapidly in recent years. However, the frequency bands suitable for current technologies have already been exploited; thus, the resources are limited. Moreover, the radio environment becomes more unpredictable because of two reasons. First, not only the volume but also the type of traffic is increasing, making the usage of the radio resource more complex. Second, distributed wireless networks such as the IEEE 802.11 wireless local area networks (WLANs) are widely deployed, and the inner- and inter-system interactions cannot be predicted easily.

Cognitive radio technologies [3, 4] have recently been developed to improve the radio resource usage of wireless networks under such situations. The basic concept of cognitive radio technology is the adaptation of the behavior of wireless systems through the recognition and learning of the radio environment. Cognitive radio systems observe and recognize the wireless network environment, make reconfiguration decisions, and apply the corresponding action to reconfigure the network. Using this approach, various radio parameters can be optimized through appropriate actions.

However, wireless systems have recently become increasingly complex. Various physical layer techniques such as code division multiplexing, frequency hopping, orthogonal frequency division multiplexing (OFDM), and Multiple-Input-Multiple-Output (MIMO) technology have been developed. Channel access techniques in MAC layer such as time division multiple access, frequency division multiple access, carrier sense multiple access with collision avoidance (CSMA/CA) have been developed. Higher layer protocol such as IP, and TCP or UDP are also used for wireless communication. Each layer technique has a number of parameters, and modern wireless systems are equipped with various combinations of those techniques. It means that the relations among radio variables and system performance are further complicated. It makes general cross-layer modeling of wireless systems difficult. Consequently, the optimization of whole wireless systems through cross-layer modeling cannot be realized.

One of the solutions for the above-mentioned issues is the machine learning technology. Machine learning is data-driven modeling that can provide a predictable model of wireless communications. The relation between the action and performance is learned by increasing the number of samples. Thus, the complex relations among various radio parameters and network performance can be obtained, which improves the precision of decision-making for the best performance.

This section proposes a wireless system optimization method based on the cognitive cycle using machine learning. The main contribution of this research is the clarification of why and how machine learning is adopted in a cognitive cycle in the current context of wireless communication. In the section 4.1, The section 4.2 introduces machine learning technologies for wireless communication systems and some related works. In the section 4.3, the formulation of the optimization of wireless systems is discussed, followed by a description of the concept of the proposed optimization method in the section 4.4. Section 4.5 and 4.6 show an implementation of the proposed scheme:

cognitive cycle and supervised machine learning based modeling. Section 4.7 shows an example of application in IEEE 802.11 WLAN, conducting the verification of the proposed method through computer simulation and experimental testbed. Section 4.8 shows another example of application in space communication, conducting the verification of the proposed method through computer emulations. Section 4.9 gives the summary.

4.1 Cognitive radio

The concept of cognitive radio was first proposed by J. Mitola [3]. Cognitive radio is described as an intelligent radio that can learn from its past experience and autonomously decide its actions suitable for radio environments and needs for communication. The proposed cognitive cycle [3, 5] is a feedback cycle of observation, learning, decision, and action. S. Haykin proposed a more concrete process of cognitive radio in [4] from an engineering perspective. He addressed the following fundamental tasks for a cognitive radio: radio-scene analysis, channel-state estimation, transmit-power control, and dynamic spectrum management. Wireless network nodes can change the radio parameters of transmission and reception in order to avoid interference among users and improve communication quality.

In general, wireless communication needs learning as a means of establishing wireless links and satisfying the communication qualities. For example, radio frequency (RF) module controls the coding rate based on the received signal strength indicator (RSSI) to reduce error probability of wireless link. It means that the RF module learn the relationship between inputs (RSSI, coding rate) and output (link quality). In cognitive radio networks, cognitive engine should determine and coordinate the actions of the cognitive radio based on learning of the environment. The relationship of inputs and outputs becomes more complicated in cognitive radio networks due to its flexibilities such as software-defined radios. The cognitive radio can control various parameters: frequency, channel, coding rate, transmission power, etc.

The relationship between these parameters and the performance of wireless communication is hardly formulated. Thus, the machine learning technologies, which can learn the complex, non-linear relationship among various information would be the solution.

4.2 Machine learning for wireless system

Technologies for next-generation wireless networks, such as 5G, are one of the major topics in the field of wireless communication today. In [26], discussions about the possibilities of machine learning technologies for the next-generation 5G network are given. Supervised learning techniques can be used to support channel state estimation in MIMO systems. Unsupervised learning for cell clustering, especially in heterogeneous networks, and reinforcement learning for the decision-making process of mobile users are also suggested. Authors of [27] have discussed Autonomic Communications in future software-driven networks. In particular, they suggested the potential of machine learning in network optimization and the needs to redesign more decentralized concepts.

In the next-generation wireless networks, networks become heterogeneous. There is a discussion of licensed shared access (LSA) [28]- [30]: 5G network nodes can use not only licensed spectrum but also unlicensed bands. S. Haykin discussed the comprehensive function of a cognitive dynamic system to organize the communications using both licensed and unlicensed bands [31]. The need for dynamic spectrum management by a cognitive dynamic system in 5G was discussed. In [32], the authors analyzed the performance optimization of heterogeneous cognitive wireless networks. A typical optimization problem of load balancing was analyzed in both centralized and decentralized cases. In [33], the authors introduced machine learning in mobile terminals in order to optimize the aggregation method for IEEE 1900.4 [34] heterogeneous wireless networks and maximize throughput.

4.3 Cross Layer Modeling of Wireless System

Several studies have attempted to understand the relations among various variables and performance to optimize wireless networks [35]- [44]. These researches generally focused on a certain layer performance such as channel capacity in physical layer and throughput in MAC layer, and does not cover higher layer application throughputs. For an example of the optimization of the wireless network capacity, we refer to the resource allocation problem in [40]. In principle, assuming ideal link adaptation, the formulation of the sum capacity of a multicell wireless network is expressed as

$$C(\boldsymbol{U}, \boldsymbol{P}) = \frac{1}{N} \sum_{n=1}^{N} \log(1 + \Gamma([\boldsymbol{U}]_n, \boldsymbol{P})),$$

where N is the total number of cells, Γ is the signal to interference and noise ratio (SINR) at the receiver, $\boldsymbol{U}$ is the set of users simultaneously scheduled across all cells, $[\boldsymbol{U}]_n$ is the users in cell n, and $\boldsymbol{P}$ is the transmit power of the scheduled users. Then, the capacity optimization problem by resource allocation is formulated as

$$\arg\max_{\boldsymbol{U},\boldsymbol{P}} C(\boldsymbol{U}, \boldsymbol{P}). \tag{4.1}$$

As referred in [40], this problem is nonconvex, so the solution is not straightforward; still this equation represents the fundamental relations among radio variables and system performance.

For another example, in [44], the optimization problem of cooperative sensing in cognitive radio networks was analyzed. This is a sensing-throughput tradeoff problem: a strict sensing policy minimizes the possibilities of interference to the primary user though the opportunities to gain more throughput would be missed, and *vice versa.* The achievable MAC layer throughputs of the secondary users R can be given as

$$R(\tau, k, \epsilon) = C_0 P(H_0) \left(1 - \frac{\tau}{T}\right) (1 - \mathbb{P}_f(\tau, k, \epsilon)),$$

where τ is the sensing time, T is the total frame time (including sensing time τ), k is the number of sensing results of sensor nodes ($1 \le k \le N$, N is the total number of sensor nodes), and ϵ is the threshold parameter of the energy detector at the sensor node. C_0 is the ideal throughput of the secondary users if the primary user is always absent, $P(H_0)$ is the probability of the primary user being absent in the channel, and $\mathbb{P}_f$ is the probability of false alarm. Focusing on the maximization of the secondary users' throughput, i.e., the minimum probability of detection of the primary user is assumed, the sensing threshold ϵ can be given by the function of τ, k, and received signal-power-to-noise ratio (SNR). Under this condition, the optimization of the throughput of secondary users is formulated as

$$\arg\max_{\tau,k} R(\tau, k). \tag{4.2}$$

Since the throughput depends on the probabilities of false alarm and detection, which depend on SNR, equation (4.2) can be expressed as a function of τ, k, and SNR. This formulation was examined by computer simulation and optimal values of τ and k for a given SNR were obtained.

The formulations of optimization problems (4.1) and (4.2) can be generalized as follows: let the radio parameters be p (such as $\boldsymbol{U}, \boldsymbol{P}$, or τ, k), the observed radio environment be z (such as SINR or SNR), and the system performance be y (such as capacity or throughput). Then, they can be formulated as $y = f(p, z)$, where f represents the relations among radio parameters, environment, and performance. Then, the optimization problem is formulated as

$$\arg\max_{\boldsymbol{p}} E(\boldsymbol{y}) = \arg\max_{\boldsymbol{p}} E(f(\boldsymbol{p}, \boldsymbol{z})), \tag{4.3}$$

where $E(\boldsymbol{y})$ is the utility function of throughputs, for example, the summation of the expected throughput of each node. By solving the above equation, the optimal set of parameters ($\boldsymbol{p}$) required to maximize the network performance is obtained. This can be done if the relation between the inputs and output is mathematically described.

In recent wireless systems, however, the situation has become more complicated. As mentioned above, modern wireless systems are equipped with various technologies in each layer. Some systems transmit signals on a single carrier with frequency hopping, and others on a multicarrier with OFDM. The channel access of one protocol is TDMA, and others' is CSMA/CA. In general, applications of wireless communication use higher layer protocol such as IP, and TCP or UDP. Therefore, we need to consider various observables $\boldsymbol{z}$ and parameters $\boldsymbol{p}$. Moreover, the relations among these variables and network performance are hardly known. Consequently, the mathematical formulation of function f cannot be realized.

Machine learning technologies, which have the fundamental characteristics of data-driven modeling, are the aid of this difficulty. By using them, the hidden and complex relations among various wireless observables and parameters and network performance can be obtained. We propose a generalized cross-layer modeling of wireless system performance using machine learning. In the proposed modeling, $\boldsymbol{E}(\boldsymbol{y})$ can denote utility of whole system performance including application. $\boldsymbol{p}$ denotes various layers' parameter, $\boldsymbol{z}$ denotes various observables. The optimization method using the proposed modeling is described in the next subsection.

4.4 Optimization based on cognitive cycle

Fig.4.1 depicts the concept of the proposed wireless system optimization method using machine learning. It is based on cognitive cycle, as described below.

4.4.1 Measurement of environment and performance

The observables of environment $\boldsymbol{z}$ are collected, which include not only the radio status but also MAC statistics, or higher layer statistics. As for $\boldsymbol{p}$, various parameters of the wireless node or network are considered. Besides these variables, network performance y is observed. They are a set of samples,

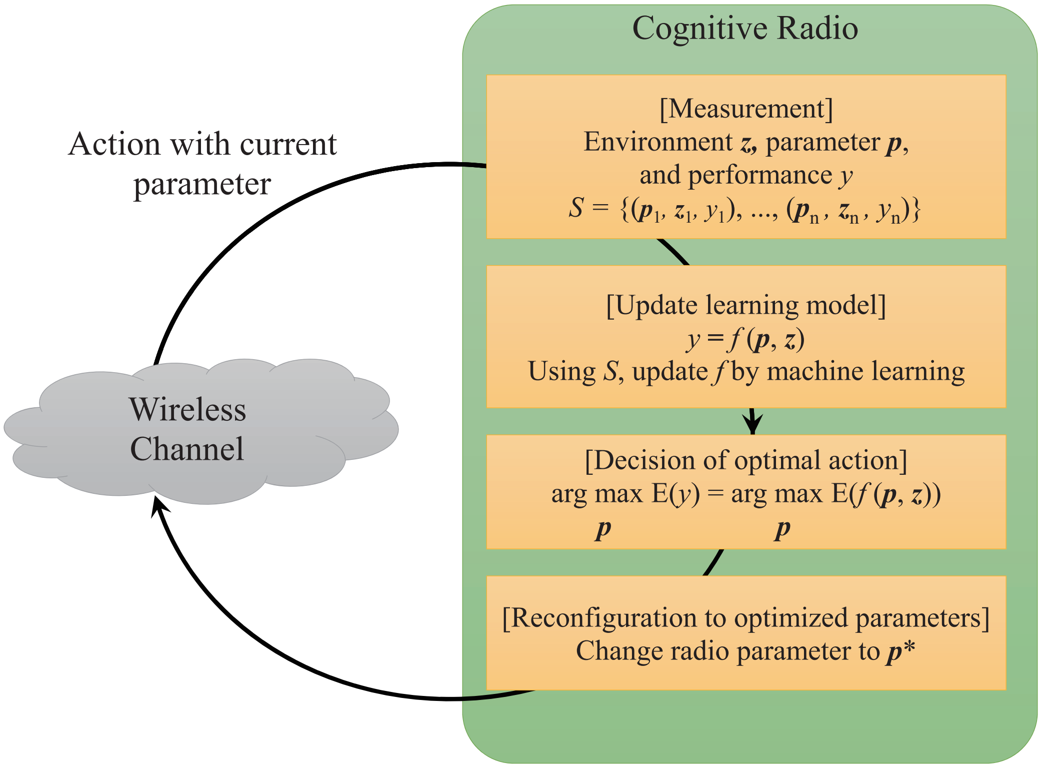


Figure 4.1: The proposed method based on cognitive cycle using machine learning.

S, for a machine learning algorithm:

$$S = \{(\boldsymbol{p_1}, z_1, y_1), (\boldsymbol{p_2}, z_2, y_2), \ldots, (\boldsymbol{p_n}, z_n, y_n)\}.$$

4.4.2 Update learning model

Using S, the cognitive engine builds and update the model f by machine learning:

$$y = f(\boldsymbol{p}, z). \tag{4.4}$$

The updating manner depends on the type of algorithm. For supervised learning, it uses S as training data, and for unsupervised learning, it uses S for clustering, or dimension reduction.

4.4.3 Decision of optimal action

By solving the optimization problem (4.3), a cognitive engine decides an optimal action to adopt the current situation. The solution of (4.3), $\boldsymbol{p}^*$, yields the optimal parameters for communication entities.

4.4.4 Reconfiguration to optimized parameters

After deciding the optimal action, to use parameter $\boldsymbol{p}^*$, the cognitive engine starts to reconfigure the wireless network. Necessary information is sent to communication entities. Fig.4.2 shows the whole system concept.

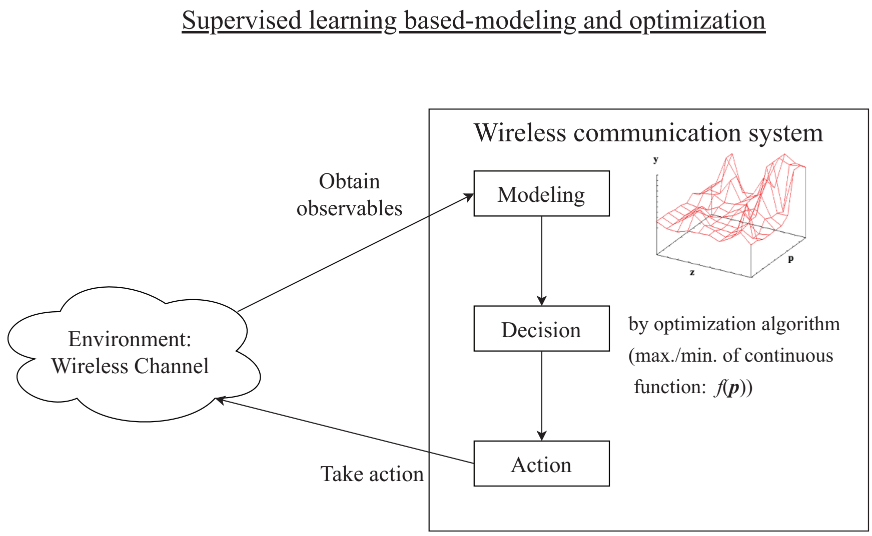


Figure 4.2: The proposed supervised learning based modeling and optimization based decisioning scheme.

In the next section, we will describe a wireless network sample for introducing and evaluating our proposed method.

4.5 Implementation and Evaluation

4.5.1 Application to IEEE 802.11 WLAN

In this section, we evaluate the proposed optimization method by applying it to the IEEE 802.11 WLAN. As an example of an optimization scenario, we consider the parameter optimization of the IEEE 802.11 stations (STAs) operated in the infrastructure mode. Each STA and cognitive controller connected to access points (APs) has functions of a cognitive engine described in the previous section and runs the cognitive cycle as mentioned below.

4.5.2 Measurement of environment and performance

Each STA measures wireless environment $\boldsymbol{z}$, obtains the current radio parameter $\boldsymbol{p}$, and performance y, and then adds a sample to $\boldsymbol{S}$. $\boldsymbol{z}$ includes the radio status at the STA, such as the received signal strength indicator (RSSI) and wireless link quality. $\boldsymbol{p}$ includes wireless parameters such as transmit power, operating channel, and address of connecting AP. The uplink or downlink throughput is considered as the performance index y.

4.5.3 Update learning model

The cognitive engine in the STA updates the learning model f by using $\boldsymbol{S}$. We consider supervised learning for the evaluation. The cognitive engine builds a model that represents the relations among $\boldsymbol{z}$, $\boldsymbol{p}$, and y from training samples $\boldsymbol{s}$, and then sends the information of the model to the cognitive controller.

4.5.4 Decision of optimal action

The cognitive controller solves the optimization problem (4.3) by using the information of the model from STAs, and obtains the optimal parameters $\boldsymbol{p}^*$ of STAs. The information of the optimal parameters is sent to STAs to reconfigure the network.

4.5.5 Reconfiguration to optimized parameters

The STA changes its wireless parameters according to $\boldsymbol{p}^*$, and then continues the cycle starting from the measurement of environment and performance.

4.6 Implementation of learning and optimization

We use support vector regression (SVR) as a learning algorithm, similar to [33]. SVR is an analog output version of support vector machines (SVMs) [45]. In SVR, the estimation function f can be expressed as [46]

$$f(\boldsymbol{x}) = \sum_{i=1}^{l}(\alpha_i' - \alpha_i)K(\boldsymbol{x}, \boldsymbol{x}_i) + b, \tag{4.5}$$

where l is the number of training samples, $\boldsymbol{x}_i$ is the input of the training samples ($\boldsymbol{p}$ and $\boldsymbol{z}$), $\boldsymbol{x}$ is an unknown input set for the learning algorithm, and K is a kernel function. α_i, α_i', and b are unknown parameters obtained by the optimization technique proposed in [46], using training samples $\boldsymbol{p}$, $\boldsymbol{z}$, and y.

We formulate the optimization problem (4.3) in the evaluation as below:

$$\arg\max_{\boldsymbol{p}} \sum_{n=1}^{N} \log\left(1 + f(\boldsymbol{p}_n, \boldsymbol{z}_n)\right), \tag{4.6}$$

where N is the number of STAs, $\boldsymbol{p}_n$ is the possible parameter set for STA-n, $\boldsymbol{z}_n$ is the current measured quality of the radio environment at STA-n, and $f(\boldsymbol{p}_n, \boldsymbol{z}_n)$ is the estimated throughput of STA-n obtained using the throughput model described above. Here, we use the logarithmic utility function of throughput considering fairness among STAs, where STAs with lower throughput have relatively larger gains for the objective function than those with higher throughput.

4.7 Application for IEEE 802.11 devices

We implement the method for the IEEE 802.11 WLAN devices. The experiments are coordinated in our university laboratory working space [49].

The IEEE 802.11 WLAN APs and STAs are operated in the 2.4 GHz ISM band. Laptop PCs with Ubuntu 14.04 are used as both STAs and APs. In each cognitive cycle, the STA observes the delay and packet loss ratio through pinging, RSSI from its connecting AP using the iwconfig command, the number of packets around the STA using tcpdump command as the link quality ($\boldsymbol{z}$), and the throughput (y) using the TCP iPerf command. The STA sets the transmission power, channel number (from 1 to 13), and data rate at the physical layer (from 6 to 54 Mb/s) for the current wireless parameters ($\boldsymbol{p}$).

The STA then builds the throughput model through SVR, and sends information regarding the SVR model to its connecting AP. The AP sends it to the cognitive controller. We have setup one of the APs as the cognitive controller, which calculates the optimal set of STA parameters $\boldsymbol{p}^*$, returns the result to the AP, and then the STA obtains the result from its connecting AP. To reduce the calculation costs for solving the optimization problem, we use the particle swarm optimization (PSO) algorithm [47,48] at the cognitive controller.

In the experiment, three APs and nine STAs are operated in channels 1, 6, and 11 in IEEE 802.11g. The operating channel is fixed for each AP. The locations of all APs and STAs are fixed during the experiment. We use uplink TCP throughputs to evaluate the performance since uplink traffic generally makes radio resource usage more competitive in CSMA/CA. We also add background UDP traffic of approximately 8 Mb/s on channel 11. To verify the performance of the proposed system, the uplink throughput performance is compared with that of other algorithms, focusing on the selection of the connecting AP at the STA as follows: (A) selection by RSSI, (B) random selection, (C) selection by radio resource utilization, and (D) selection of the number of STAs as equally as possible among channels. In algorithm (A) using RSSI, the STA selects an AP with the highest RSSI. This seems to be a popular method for devices in the market. In algorithm (C), the STA selects the AP of a channel where the minimum number of packets is observed in each cycle. In each algorithm, each cycle runs for 30 s. All STAs start iPerf traffic of 2 s at the same time in each cycle. Before starting the proposed method, the STA observes the radio environment in each channel for 1 hour and utilizes it as training data.

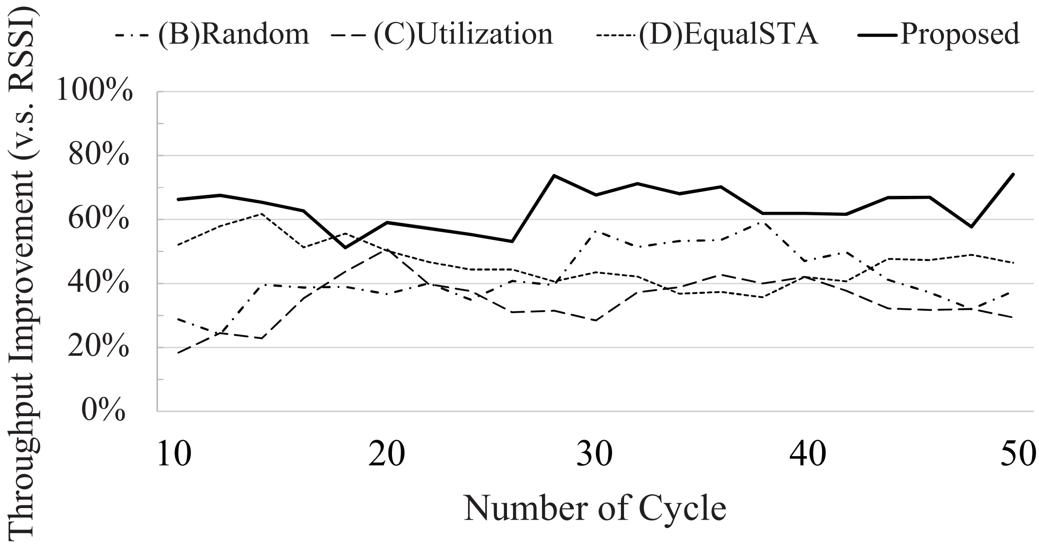


Figure 4.3: Improvement in throughput relative to (A)RSSI by time in each algorithm. Copyright(C)2020 IEICE, [302] Fig.2

Fig. 4.3 shows the moving average throughput by time for each algorithm. Throughputs are normalized by those in RSSI. The time is expressed as the number of cognitive cycle, and the throughput is averaged every 10 cycles (5 minutes). The proposed method shows greater throughput than other algorithms, indicating that the STAs can select APs effectively.

Fig. 4.4 compares the average throughput per channel among the algorithms. The utilization-based algorithm (C) shows higher throughput at channel 6, where it is detected as the most vacant channel. However, the throughputs at the other channels are much lower. This algorithm is based

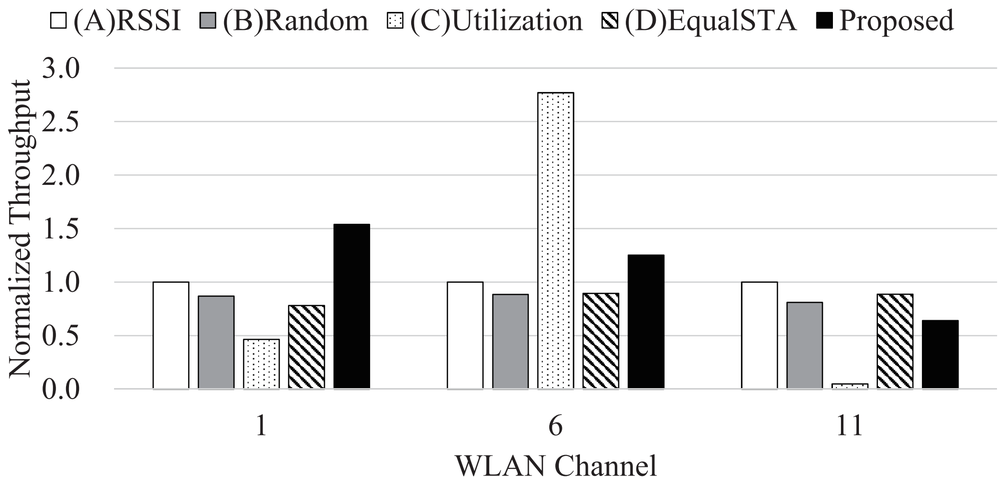


Figure 4.4: Average throughput at each channel. Throughput is normalized by those in (A)RSSI. Copyright(C)2020 IEICE, [302] Fig.2

on the observations of a wireless environment but does neither learns, nor optimizes the whole system.

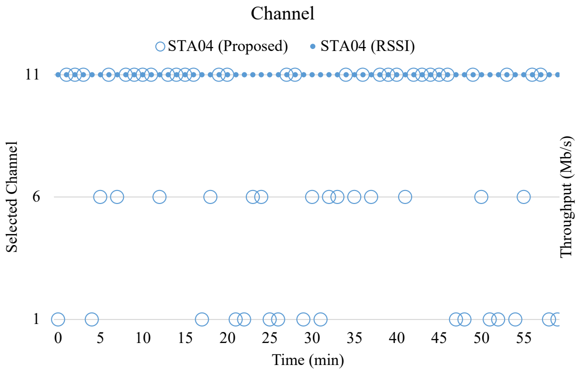
In contrast, the proposed method, which has a function of learning and optimization, shows higher throughput at channels 1 and 6 and lower throughput at channel 11, which has higher background traffic. As a whole, the proposed method can improve the network performance. These results indicate that the proposed method can build the appropriate throughput model through learning, and can select the optimized wireless parameters that improves the whole network performance.

Channel selection at each station

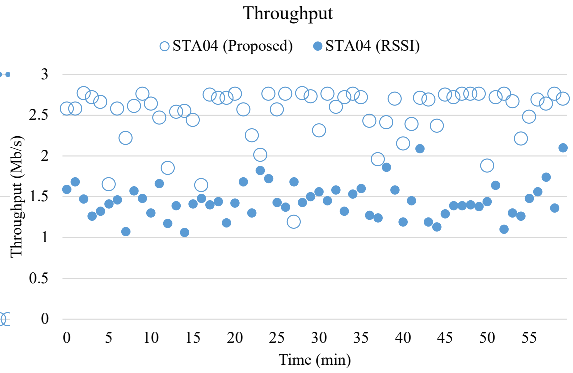
In order to investigate the detail of the results, the total system throughput increased, the channel selection and the throughput of STAs are examined here. Fig.4.5 and Fig.4.6 show the channel selections and the throughputs of STAs whose throughputs are increased with the proposed algorithm compared to those with RSSI. STA04, for example, selects channels of 1 and 6 with the proposed algorithm, whereas with RSSI based algorithm, it selects only channel of 11. The throughputs with RSSI based algorithm are lower than those with the proposed algorithm though the time.

Fig.4.7 shows the channel selections and the throughputs of STAs whose throughputs keep those levels. STA02, for example, selects channels of 6 and 11 with the proposed algorithm, whereas with RSSI based algorithm, it selects only channel of 1. The throughputs with the proposed algorithm are similar to those with RSSI based algorithm though the time, sometimes lower and sometimes higher, and even as average.

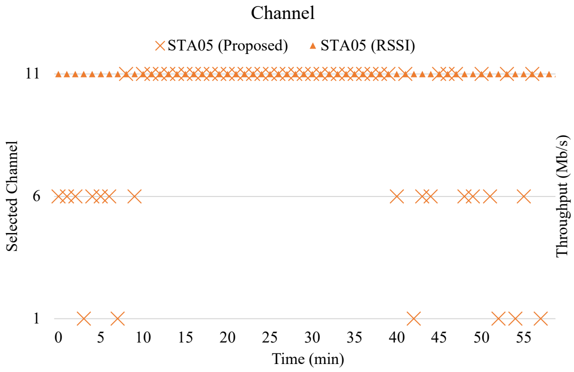
Fig.4.8 shows the channel selections and the throughputs of STAs whose throughputs keep those levels. With RSSI algorithm, these STAs select channels of 1 or 6, because RSSI to APs using these channels is higher than that of channel 11. On the other hand, with the proposed algorithm, STAs select sometimes channel 11. Due to the background traffic applied to channel 11, the throughputs in the proposed algorithm decrease. Note that the throughputs with RSSI algorithm are relatively higher than those in Fig.4.5 and Fig.4.6, where the throughputs increase. This results come from the optimization algorithm introduced in Eq.(4.6), i.e. STAs with lower throughputs and higher throughputs are optimized by this equation.



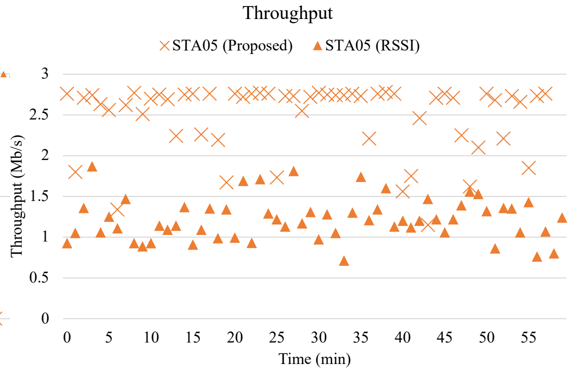
(a) Channel selection at STA04.



(b) Throughput at STA04.

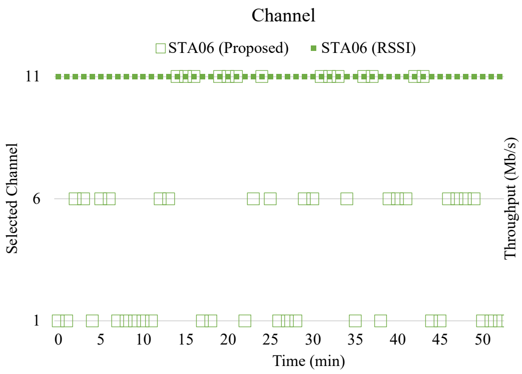


(c) Channel selection at STA05.

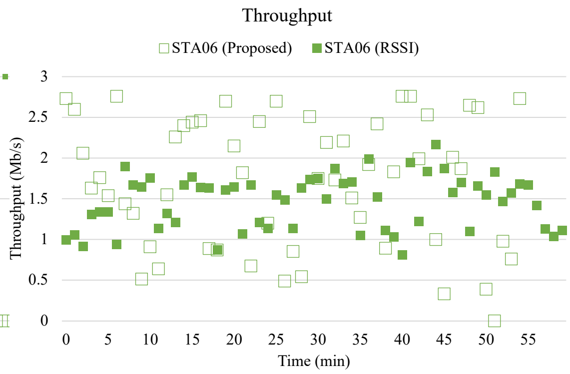


(d) Throughput at STA05.

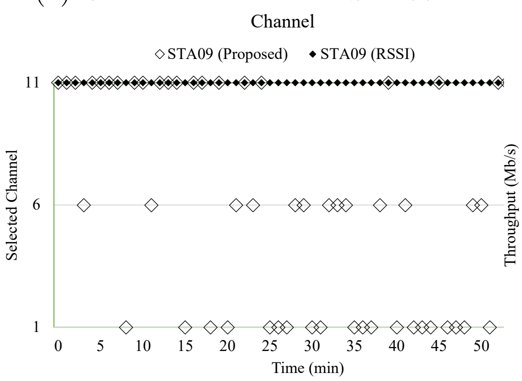
Figure 4.5: STAs where throughputs increase.



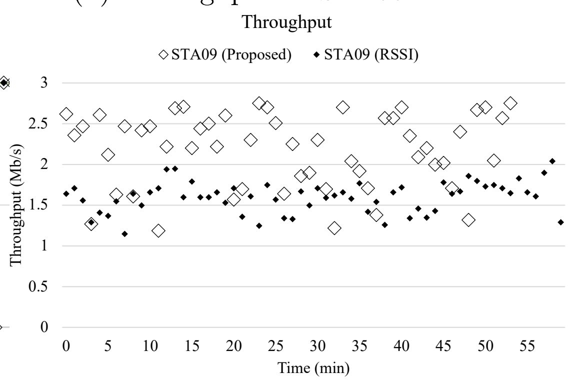
(a) Channel selection at STA06.



(b) Throughput at STA06.

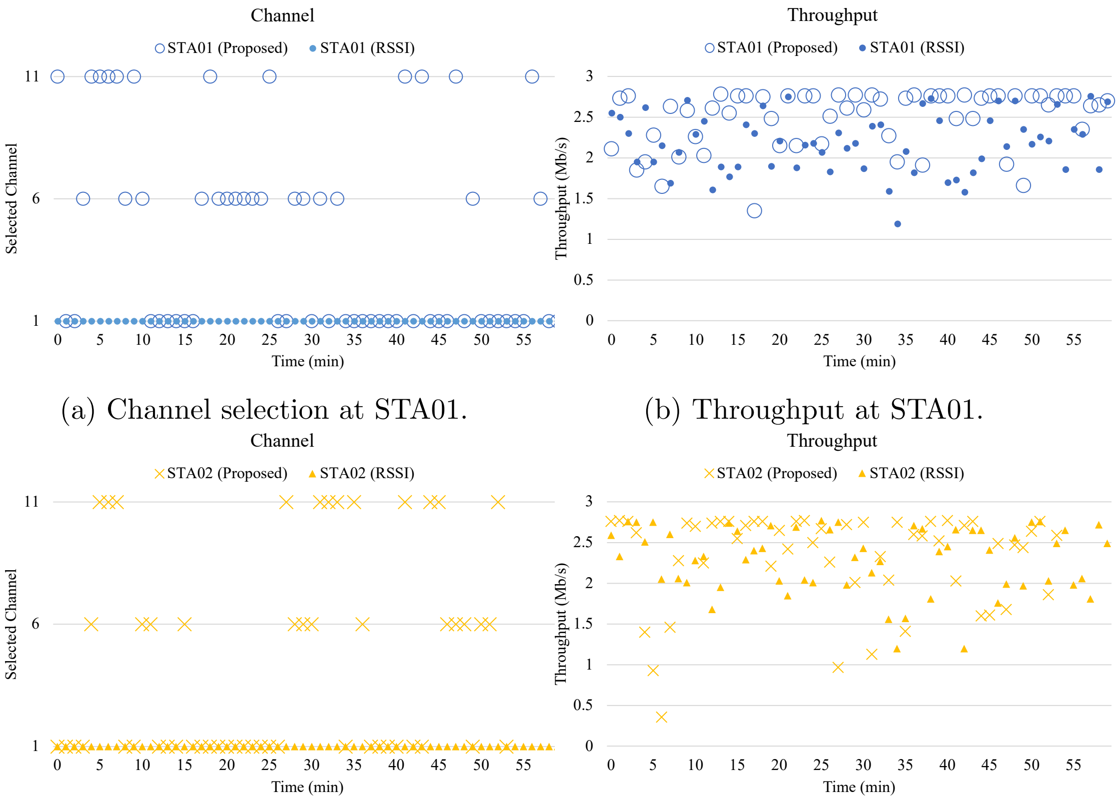


(c) Channel selection at STA09.



(d) Throughput at STA09.

Figure 4.6: STAs where throughputs increase (continued).



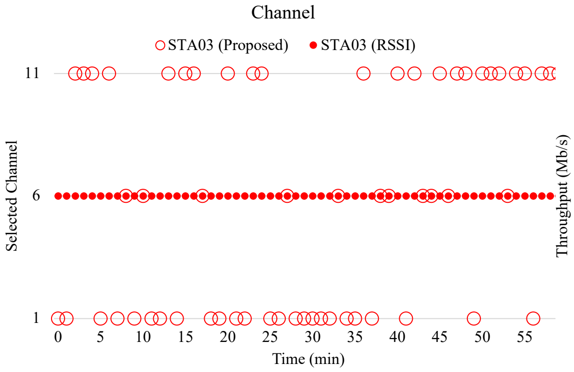
(a) Channel selection at STA01.

(b) Throughput at STA01.

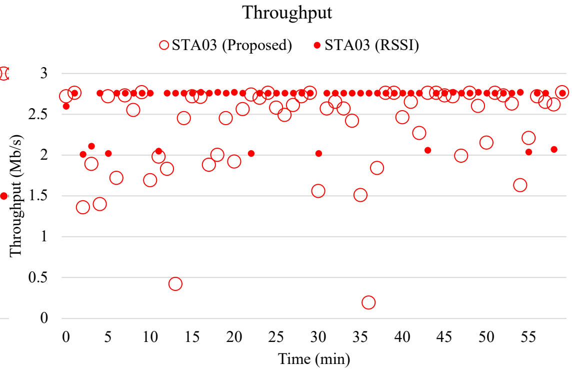
(c) Channel selection at STA02.

(d) Throughput at STA02.

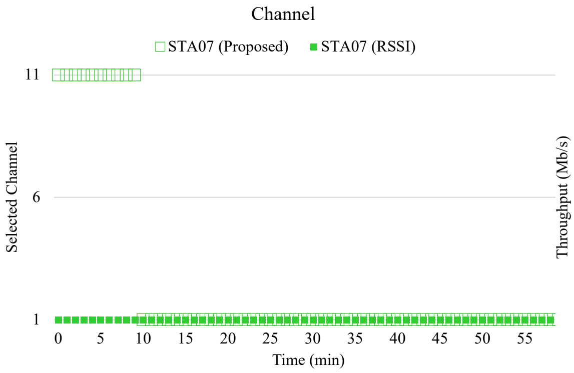
Figure 4.7: STAs where throughputs are similar.



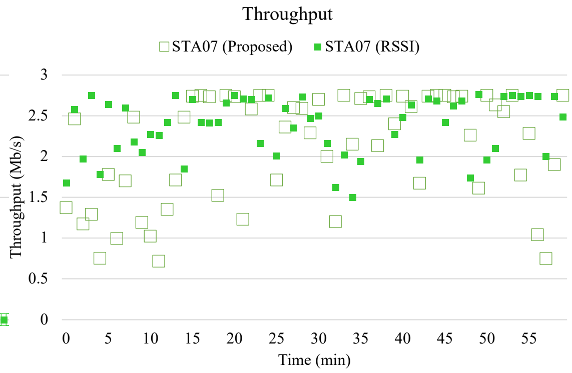
(a) Channel selection at STA03.



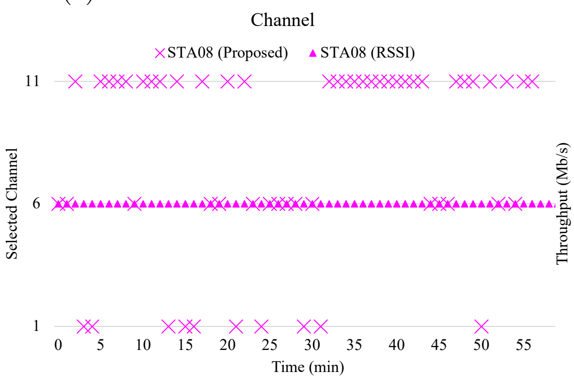
(b) Throughput at STA03.



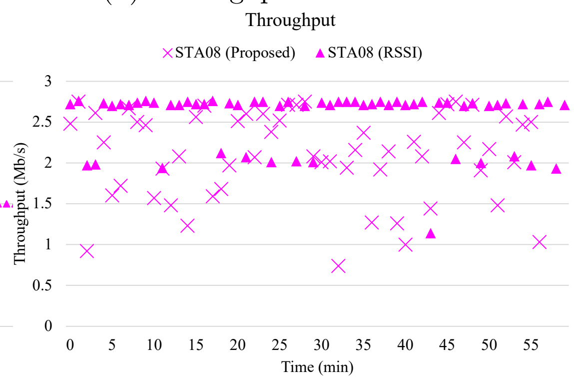
(c) Channel selection at STA07.



(d) Throughput at STA07.



(e) Channel selection at STA08.



(f) Throughput at STA08.

Figure 4.8: STAs where throughputs decrease.

4.7.1 Evaluation by computer simulation

We also conduct a computer simulation for an extended evaluation of the proposed method. The basic implementation is the same as that in the experiments already shown. The binary programs of learning and optimization are also the same as those in the experimental devices. Network simulator QualNet 7.4 [50] is used for the platform of computer simulation. The number of STAs is 21 and that of APs is 3; the operating channels are 1,6, and 11. The variables of learning sample $(\boldsymbol{p}, \boldsymbol{z}, y)$ are the same as those in the experiment conducted in the laboratory. STAs in the proposed system send uplink TCP traffic of two types of offered loads. The background traffic is generated by constant bit rate (CBR) traffic. The offered load of channel 6 is rather smaller than that of channels 1 and 11. The detailed settings of simulation are shown in Table 4.1. The main difference in settings from those of the experiments is the offered load variation of STAs in the proposed system. Similar to the experimental results, computer simulation shows the improvement by introducing the proposed method, as shown in Fig. 4.9 and Fig. 4.10. From Fig. 4.10, the proposed cognitive cycle using machine learning can optimize the choice of channel in accordance with the formulation of (4.6).

Channel selection at each station

In order to investigate the detail of the results, the total system throughput increased, the channel selection and the throughput of STAs are examined here. Fig.4.11–4.13 show the channel selections and the throughputs of STAs whose throughputs are increased with the proposed algorithm compared to those with RSSI. STA04, for example, selects channels of 1 and 6 with the proposed algorithm, whereas with RSSI based algorithm, it selects only channel of 11. The throughputs with RSSI based algorithm are lower than those with the proposed algorithm though the time.

Table 4.1: Simulation settings

Parameter	Value
Area size	20 m x 20 m
Number of ieteration	10 times
Pathloss model	Free space decay
Channel model	Additive white gaussian noise (AWGN)
Traffic in proposed system	TCP of 1.4 Mbps in 6 STAs
	TCP of 0.7 Mbps in 15 STAs
Number of background APs	3 APs in each channel
Number of background STAs	5 STAs in channel 1
	1 STA in channel 6
	7 STAs in channel 11
Background traffic	CBR of 500 Kbps/STA in channel 1
	CBR of 100 Kbps in channel 6
	CBR of 500 Kbps/STA in channel 11

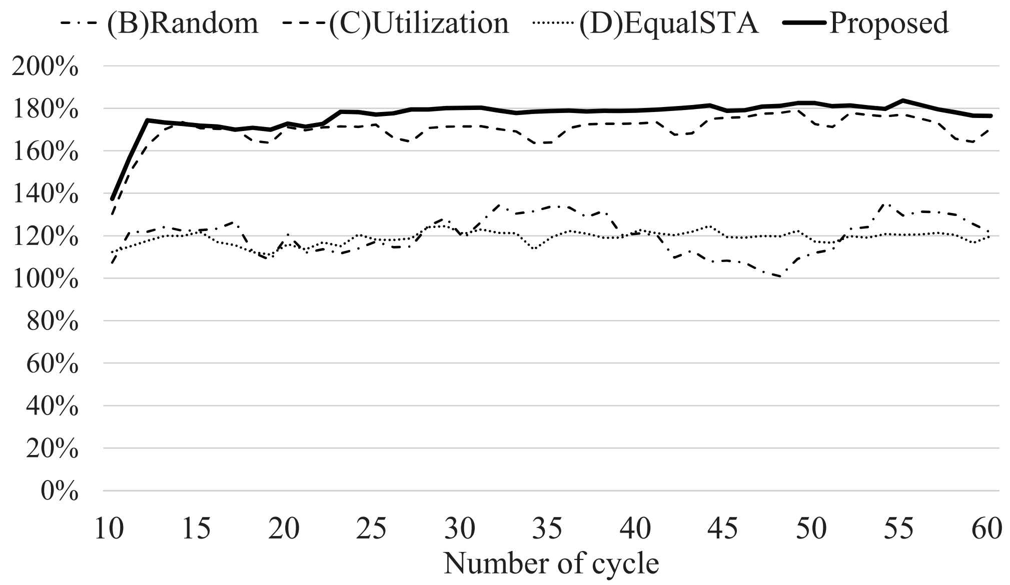


Figure 4.9: Improvement in throughput relative to (A)RSSI by time in each algorithm in computer simulation.

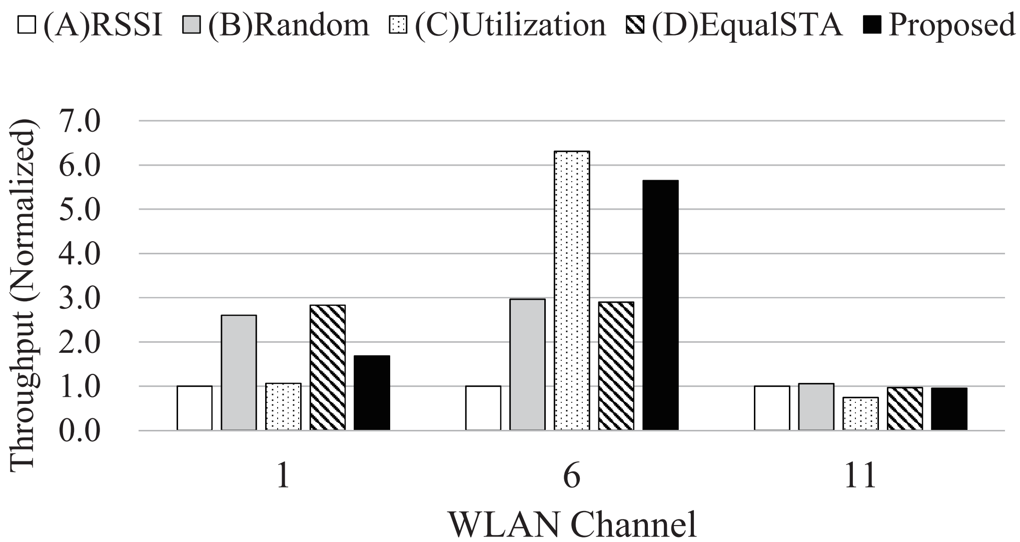
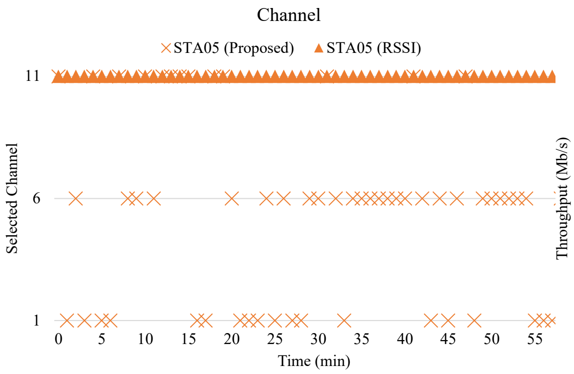
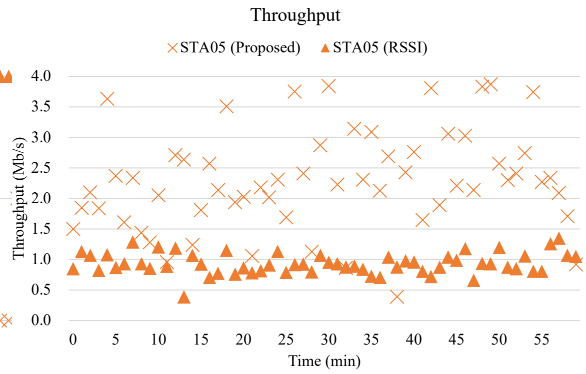


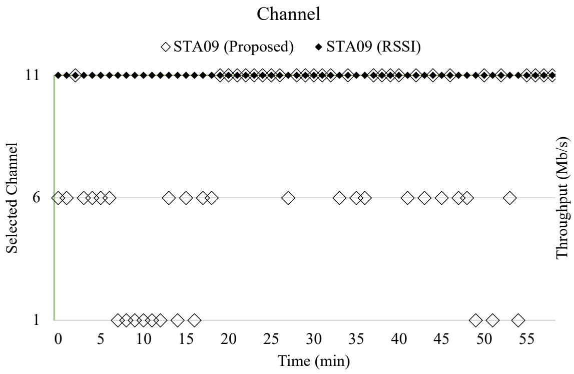
Figure 4.10: Average throughput at each channel in computer simulation. Throughput is normalized by those in (A)RSSI.



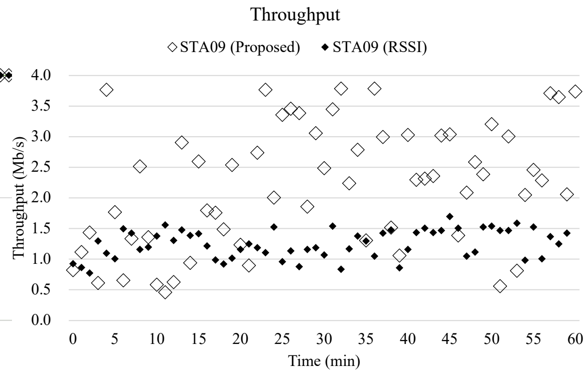
(a) Channel selection at STA05.



(b) Throughput at STA05.

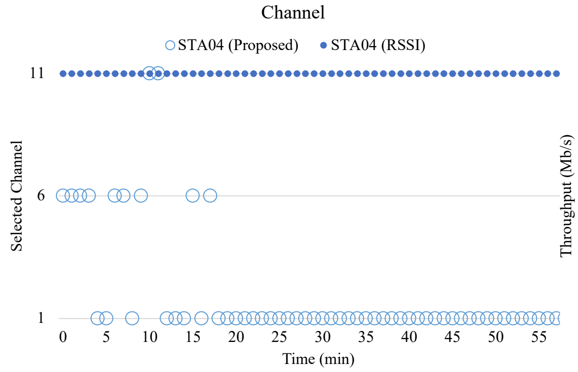


(c) Channel selection at STA09.

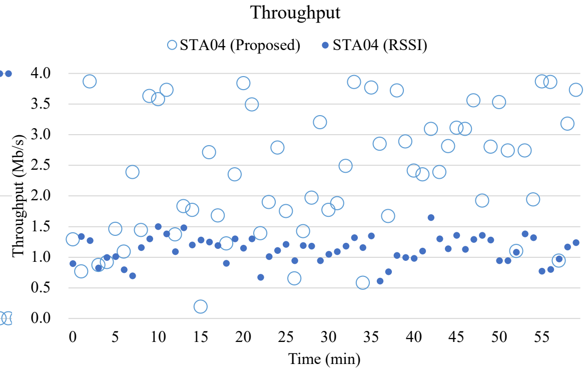


(d) Throughput at STA09.

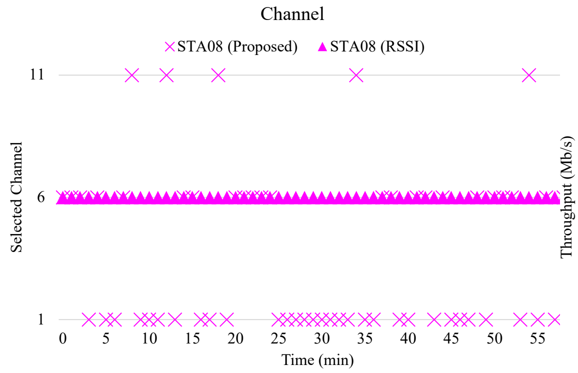
Figure 4.11: STAs where throughputs increase.



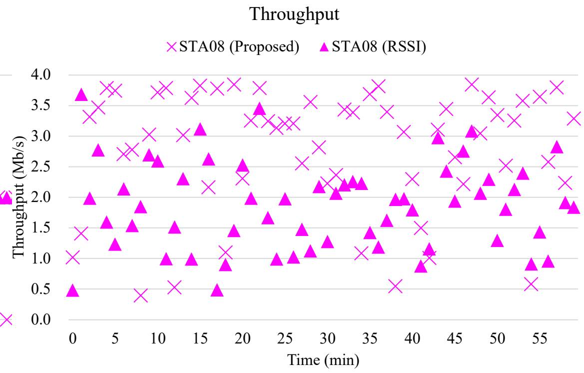
(a) Channel selection at STA04.



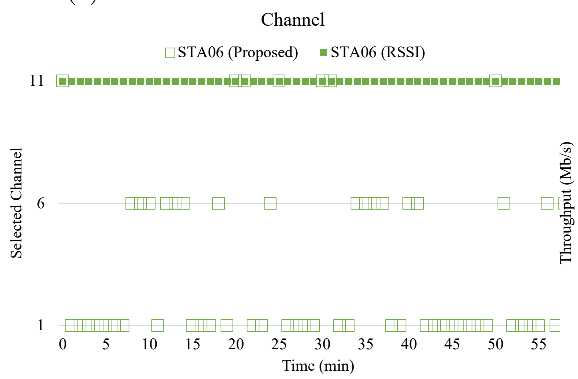
(b) Throughput at STA04.



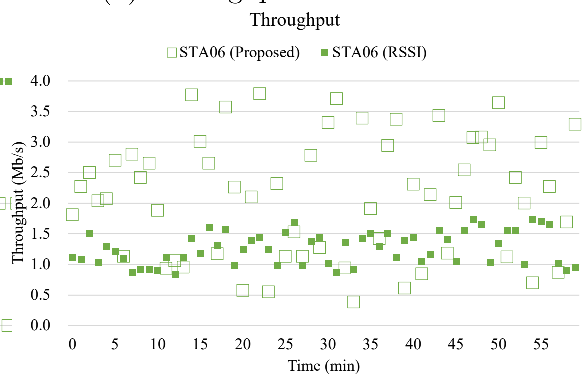
(c) Channel selection at STA08.



(d) Throughput at STA08.

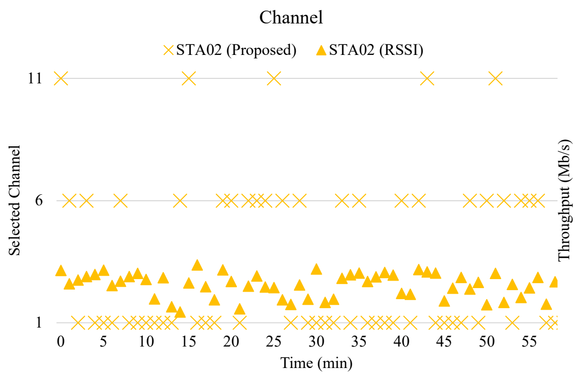


(e) Channel selection at STA06.

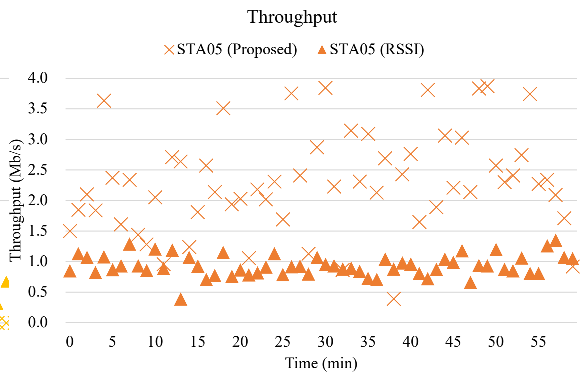


(f) Throughput at STA06.

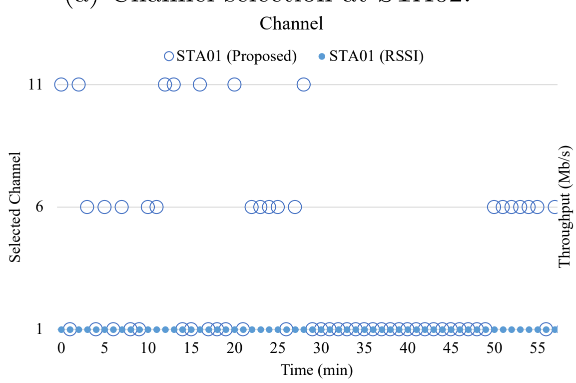
Figure 4.12: STAs where throughputs increase (continued).



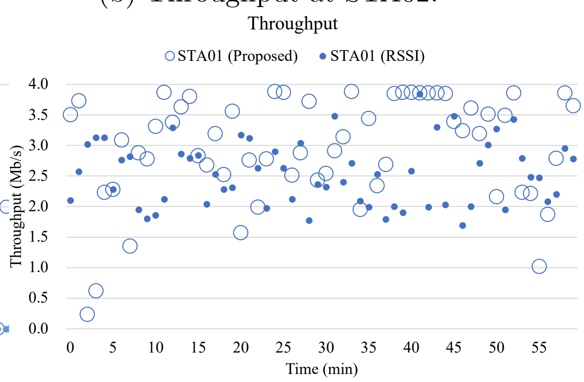
(a) Channel selection at STA02.



(b) Throughput at STA02.

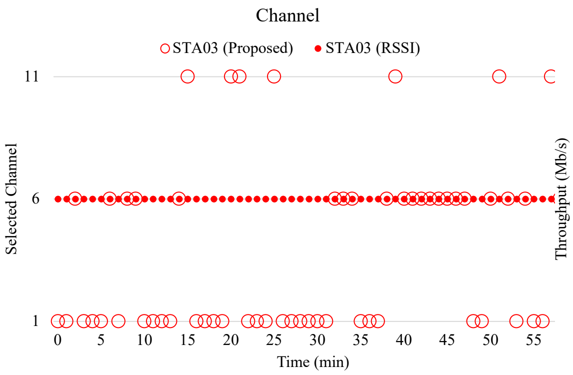


(c) Channel selection at STA01.

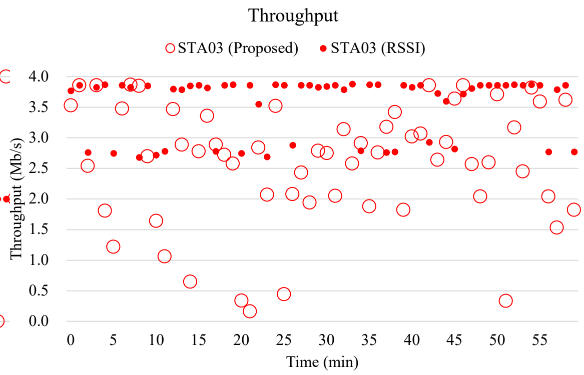


(d) Throughput at STA01.

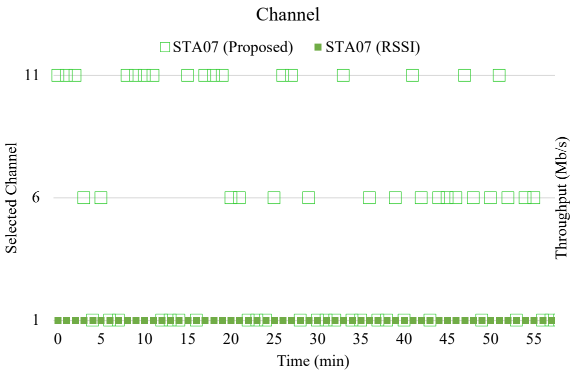
Figure 4.13: STAs where throughputs increase (continued).



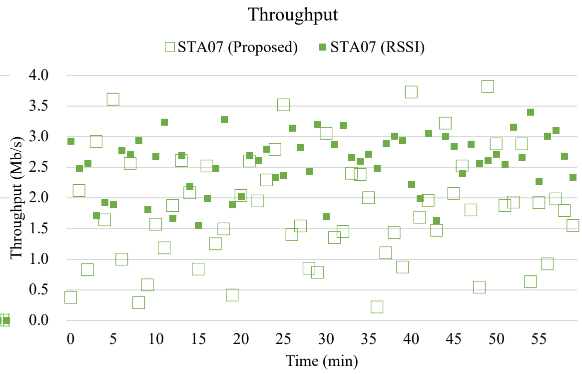
(a) Channel selection at STA03.



(b) Throughput at STA03.



(c) Channel selection at STA07.



(d) Throughput at STA07.

Figure 4.14: STAs where throughputs decrease.

4.8 Application for space communication

The propose supervised learning based optimization scheme is applied for space communication. Fig.4.15 shows an example of future wireless communication system in the space, indicated in [295].

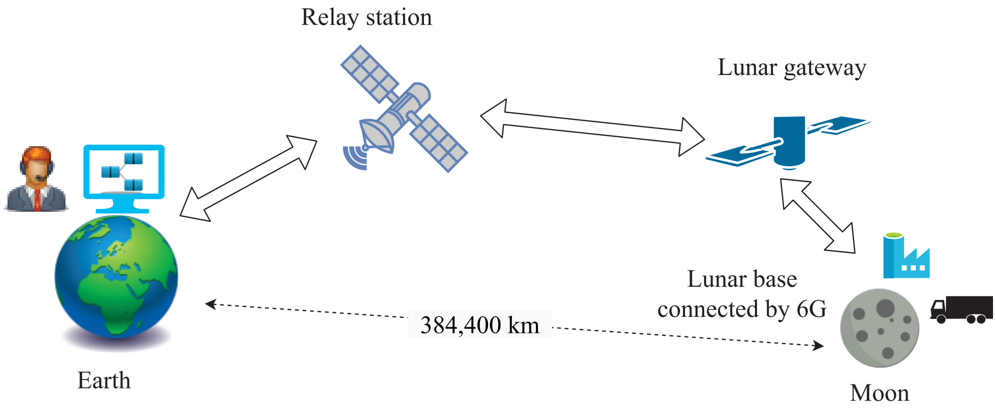


Figure 4.15: An example of future wireless communication between the Earth and the Moon. The basic concept is in [295]

One of the characteristic of space communication different from terrestrial wireless communication is large communication delay. The main component of communication delay is a propagation delay, which is shown in Table 4.2. The distance from the Earth to the Moon, for example, is 384,400 km in average, where the bidirectional wireless communication experiences at least more than 2.56 seconds. This scale of delay is to be handled at higher layer than physical or MAC layers, namely, transport layer with such as TCP. However, the physical and MAC parameters have also to be taken into consideration: modulation and coding scheme (MCS), transmission power etc. This is the similar situation that is shown in previous subsection in the application for IEEE 802.11 devices. Therefore, in this subsection, the application of proposed supervised learning based optimization scheme to the space communication is examined.

Table 4.2: The relation between distance in one way (km) and two way propagation delay (ms). The propagation media is assumed to be vacuum, i.e. the refractive index is 1.0.

Distance (one way, km)	Propagation delay (two way, ms)
1,000	6.67
5,000	33.36
10,000	66.71
30,000	200.14
50,000	333.56
100,000	667.13
200,000	1,333.26
300,000	2,001.38
400,000	2,668.51
450,000	3,002.08

4.8.1 System model

Fig.4.16 shows the system model of evaluation. The wireless communication network is simplified a direct communication between the ground station on the Earth and the satellite near the Moon. The ground station run the proposed scheme: optimal decisioning of communication parameters of from physical layer to transport layer. Here, the specifications of physical and

Table 4.3: Availabe MCS

MCS index	Modulation and coding rate
1	QPSK 1/4
2	QPSK 1/3
3	QPSK 2/5
4	QPSK 1/2
5	QPSK 3/5
6	QPSK 2/3
7	QPSK 3/4
8	8PSK 2/3
9	8PSK 3/4
10	16QAM 2/3
11	16QAM 3/4
12	32QAM 2/3
13	32QAM 3/4

MAC layer are abstract models: wireless communication nodes, including the base station on the Earth and a satellite near the Moon, transmit signals using some modulation and coding scehems (MCSs). Table 4.3 shows the MCS indexes of the evaluation. Other parameters, MTU and TCP algorithms are: MTU 500 to 1,500 Byte, and TCP algorithms are Reno, cubic, and Bottleneck Bandwidth and Round-trip propagation time (BBR) [296].

4.8.2 Implementation and Evaluation

Table 4.4 shows the algorithms for evaluation. These corresponds with right quadrants of Fig.3.2, i.e. the amount of available information is large and the decisioning is done by learning or optimization algorithm.

Table 4.4: Algorithms of modeling and decisioning using supervised learning.

ID	Modeling	Decisioning
(1)	Deep learning	Reinforcement learning: MAB with ϵ-Greedy
(2)	Support vector regression	Reinforcement learning: MAB with ϵ-Greedy
(3)	Support vector regression	Optimization (using PSO)

Desktop computer where Ubuntu 17.10 is installed is used for emulation of space communication, by using tc command. For the application traffic, image file of 10 MByte is transfered by a socket program. The duration of transferring is monitored to obtain the throughput. Parameters such as delay corresponding radio propagation in space and other processing factors, and packet error rare are variable. In order to train each alorithm, several pre-traing with random sampling parameters was conducted before the experiments and used for each algorithm to build the model. Table 4.5 shows other settings of parameters.

As a reference algorithm, a deep reinforcement learning, as previously shown in Fig.3.3 is examined through experiments. It is to evaluate the right panes of Fig.3.2.

In Fig.4.17 shows the throughput results of those algorithms with the communication distance of 390,000 km. It roughly corresponds to the distance between the Earth and the Moon (384,400 km), where the thoughput performances show the superiority of the proposed algorithm. The percentage values show the relative throughput increase or decrease to that of deep reinforcement learning (DRL) algorithm (deep learning with ϵ-Greedy algorithm). The proposed algorithm, support vector regression (SVR) modeling and optimization algorithm (using PSO), show +18 % increase of throughput. On the other hand, the through put of supervised learning modeling and reinforcement learning decision, i.e. SVR and ϵ-Greedy algorithm, show

Table 4.5: Parameters for experiments.

Parameter	Value
iterations	10 times
number of pre-traing iterations	500
commnication delay	propagation delay in space
pathloss model	free space attenuation
center frequency	14.25 GHz

lower throughput than that of DRL, which means +60 % increase by using optimization algorithm, instead of ϵ-Greedy algorithm.

Fig.4.18–Fig.4.20 show the parameters selected by each algorithm. Fig.4.18 corresponds to the algorithm (1), deep learning based modeling and reinforcement learning based decisioning with ϵ-Greedy algorithm, Fig.4.19 corresponds to the algorithm (2), SVR based modeling and reinforcement learning based decisioning with ϵ-Greedy algorithm, Fig.4.20 corresponds to the proposed algorithm (3), SVR based modeling and optimization algorithm based decisioning (with PSO).

Fig.4.22 shows the parameters selected by the algorithm (1), deep learning based modeling and reinforcement learning based decisioning with ϵ-Greedy algorithm in communication distance of 450,000 km.

Fig.4.23 shows the parameters selected by the algorithm (2), SVR based modeling and MAB algorithm based decisioning in communication distance of 450,000 km.

Fig.4.24 shows the parameters selected by the algorithm (3), SVR based modeling and optimization algorithm based decisioning (with PSO) in communication distance of 450,000 km.

4.9 Summary

In this section we proposed a wireless system optimization method based on the cognitive cycle using machine learning. We reviewed several studies of traditional wireless system optimization problem, and showed that formulations of those problems can be generalized as the relations among radio variables and system performance. Modern wireless systems are becoming increasingly complex, and the optimization of them is a challenging issue. Based on the formulation of the optimization problem, we indicated that the machine learning technology, data-driven modeling that can provide a predictable model of wireless communications, can be a solution. The proposed method was evaluated by applying it to the IEEE 802.11 WLAN. Since the proposed method is based on a comprehensive, learning-based cognitive cycle, it can be applied to various wireless system optimization problems such as an adaptive MIMO parameter selection, a frequency selection and a transmit power control in heterogeneous cellular networks, and the selection of radio interfaces in a smartphone. It is important to investigate what type of machine learning algorithm is suitable for each system. From the viewpoint of practical implementation, it is also necessary to develop optimization algorithms to solve the problem of a large scale network.

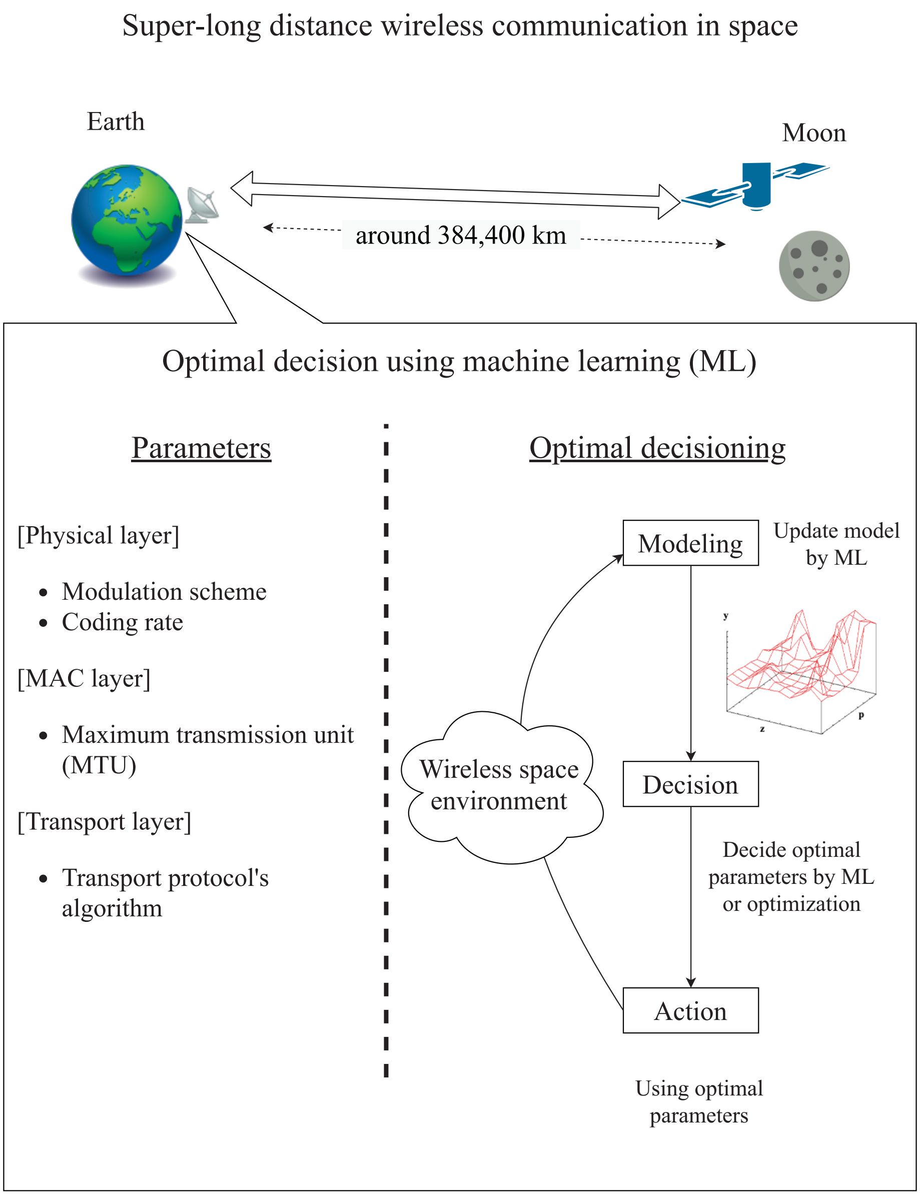


Figure 4.16: A system model to verify the proposed scheme in the future wireless communication between the Earth and the Moon as an example of proposed scheme.

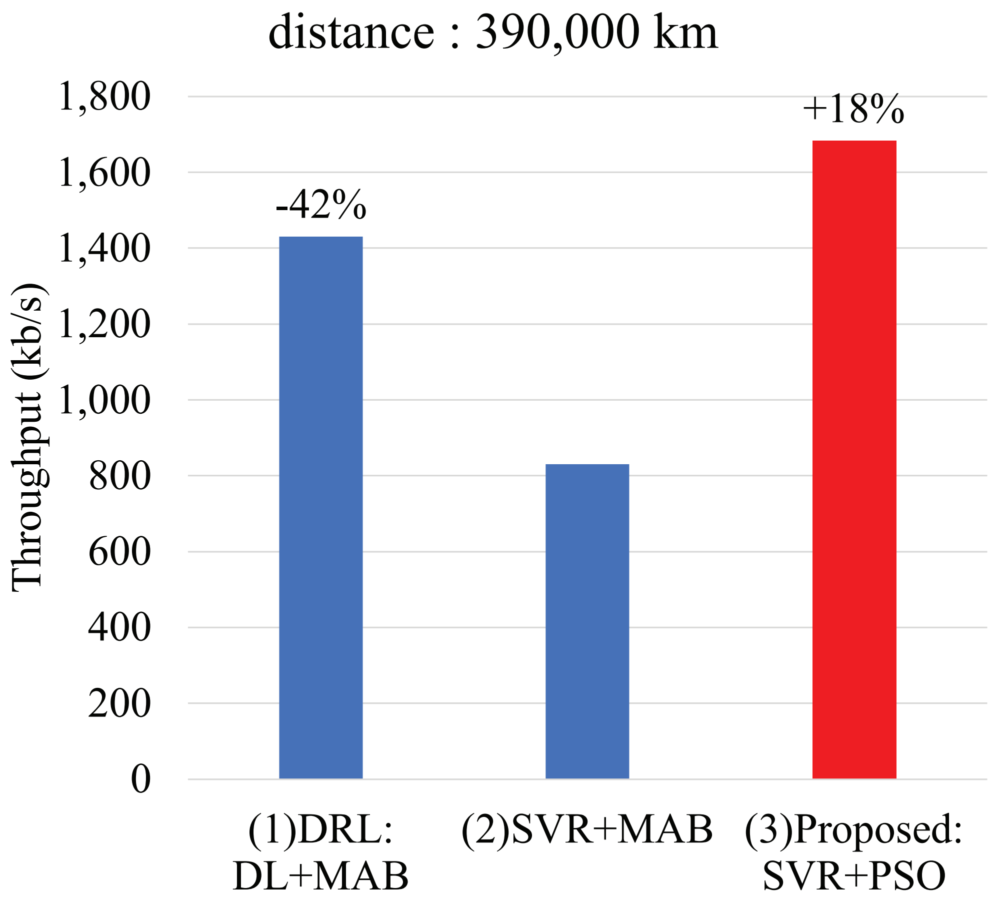
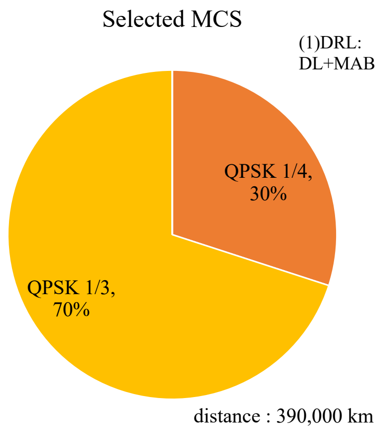
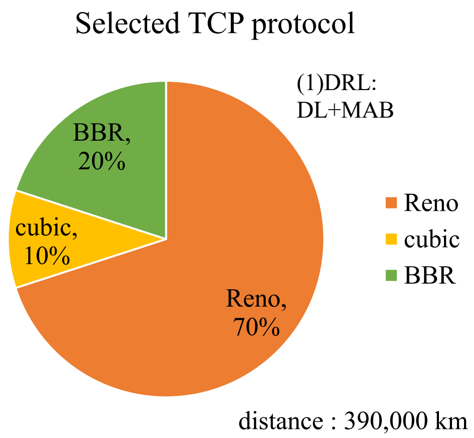


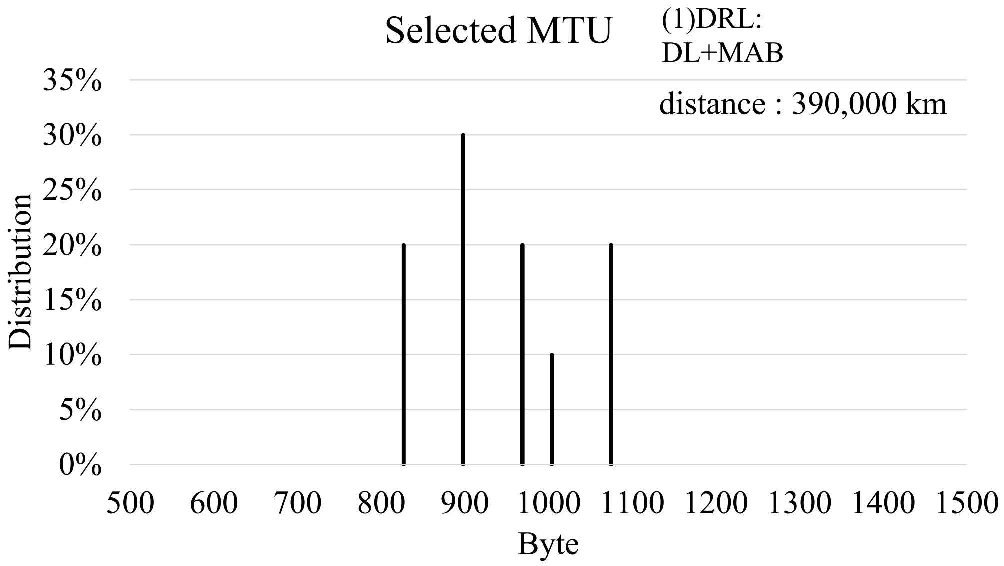
Figure 4.17: Throughput of each algorithm with communication distance 390,000 km.



(a) Selected MCS by SVR+PSO.

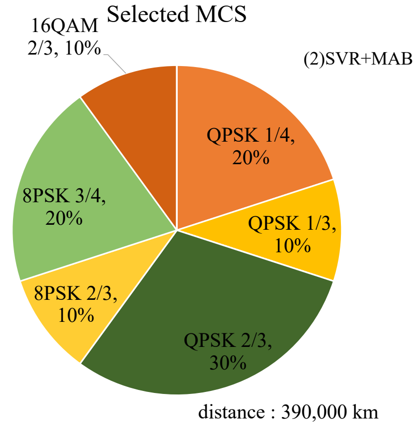


(b) Selected TCP algorithm by SVR+PSO.

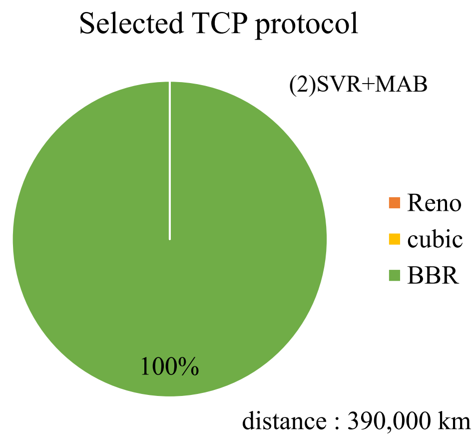


(c) Selected MTU by SVR+PSO.

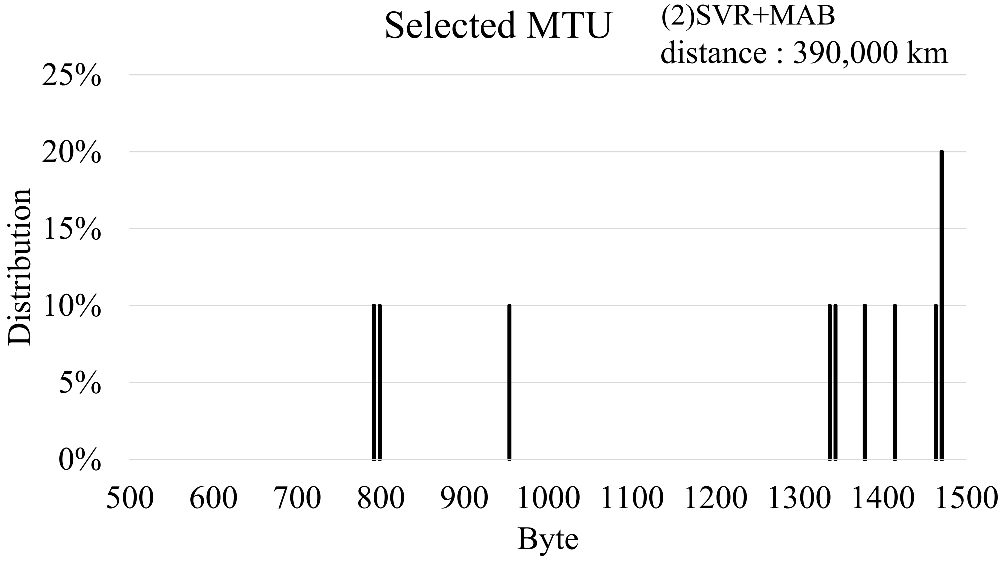
Figure 4.18: Selected parameters through iterations with algorithm (1) in 390,000 km.



(a) Selected MCS by SVR+PSO.

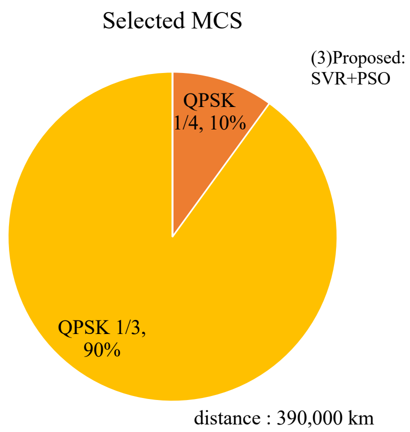


(b) Selected TCP algorithm by SVR+PSO.

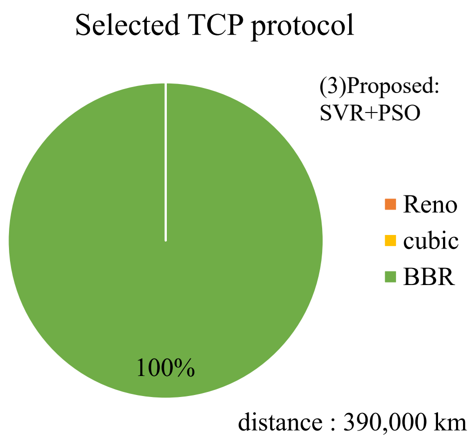


(c) Selected MTU by SVR+PSO.

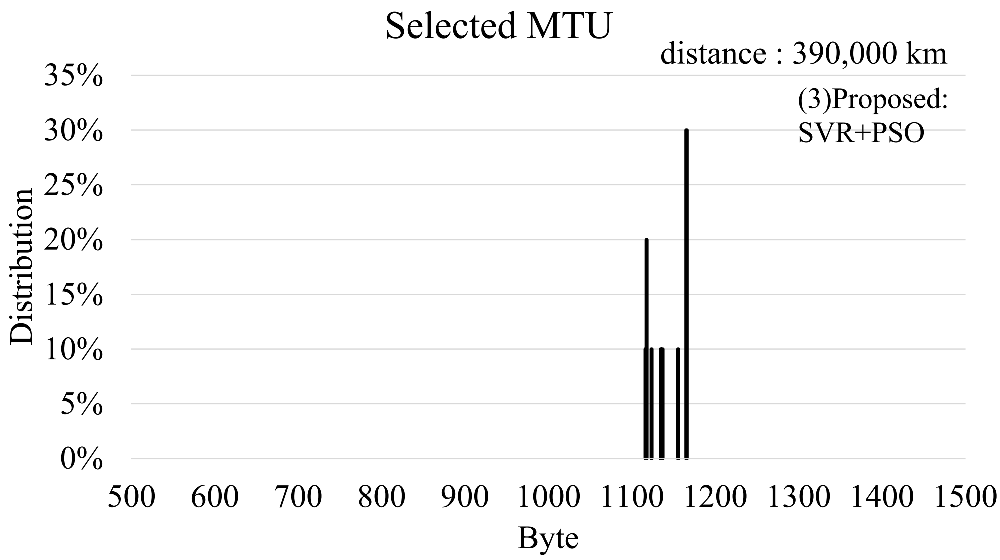
Figure 4.19: Selected parameters through iterations with algorithm (2) in 390,000 km.



(a) Selected MCS by SVR+PSO.



(b) Selected TCP algorithm by SVR+PSO.



(c) Selected MTU by SVR+PSO.

Figure 4.20: Selected parameters through iterations with algorithm (3) in 390,000 km.

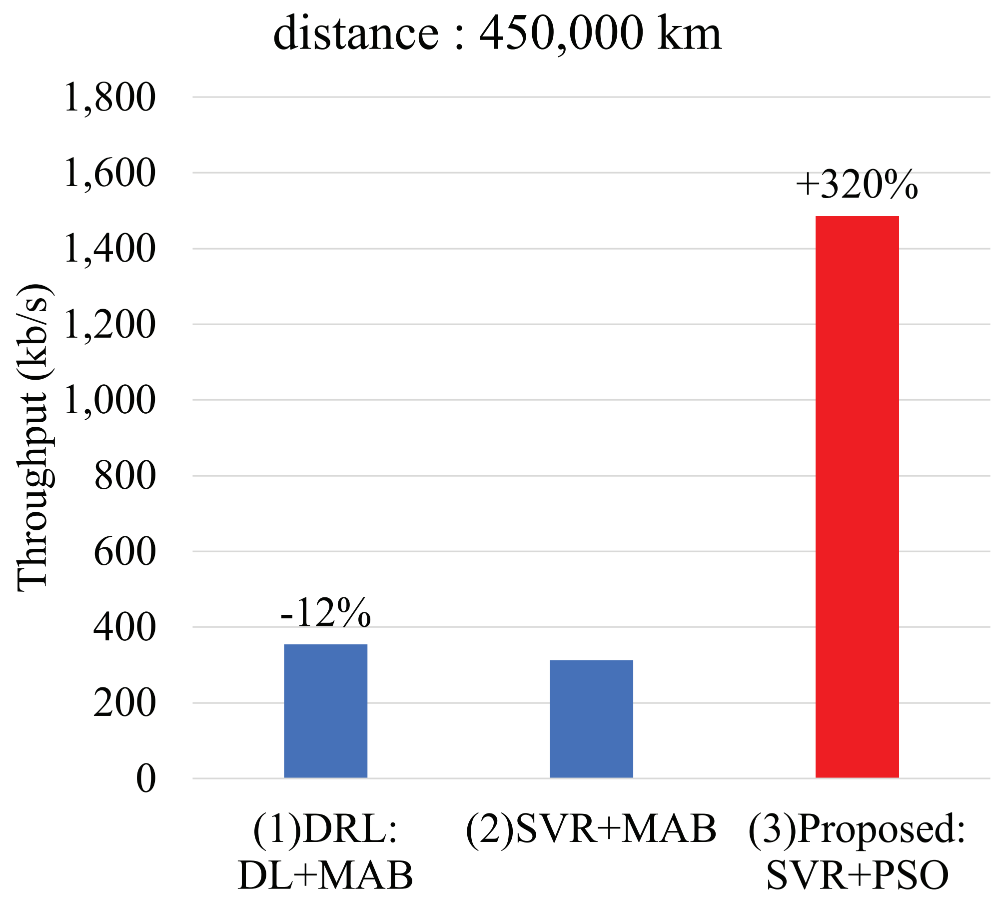
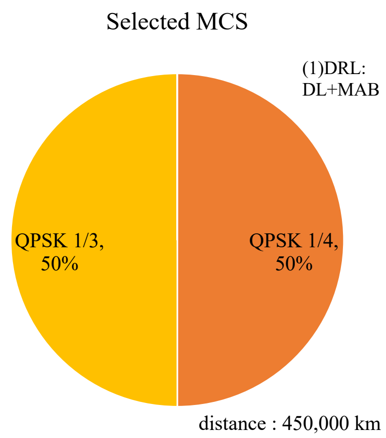
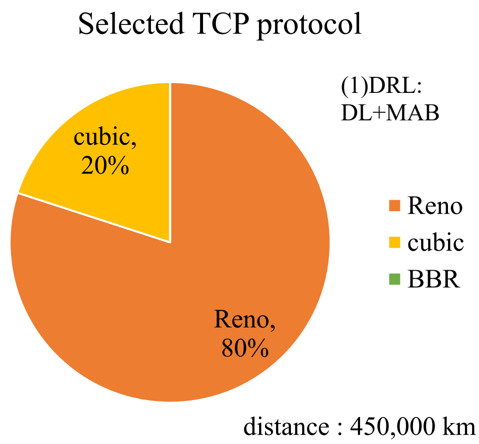


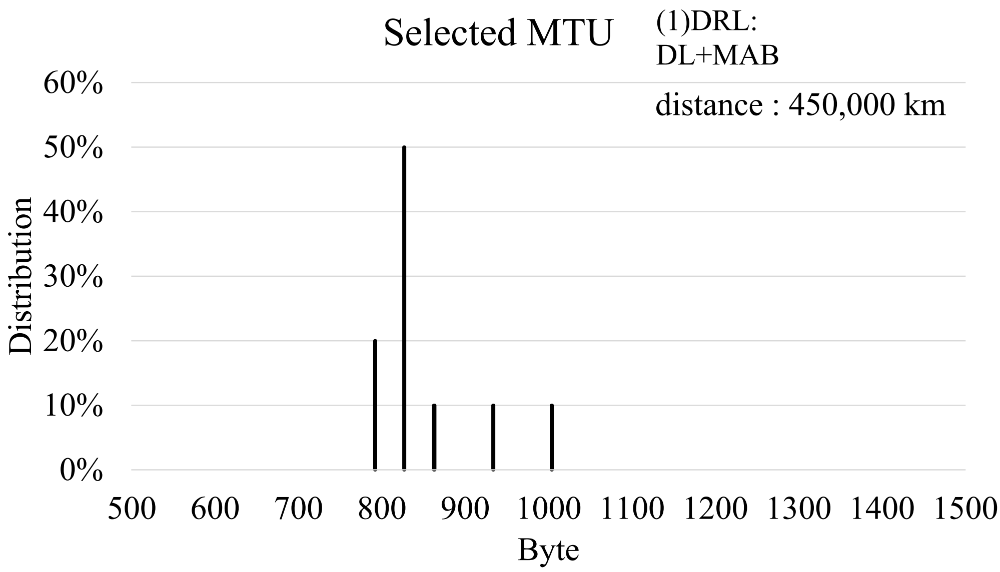
Figure 4.21: Throughput of each algorithm with communication distance 450,000 km.



(a) Selected MCS by SVR+PSO.

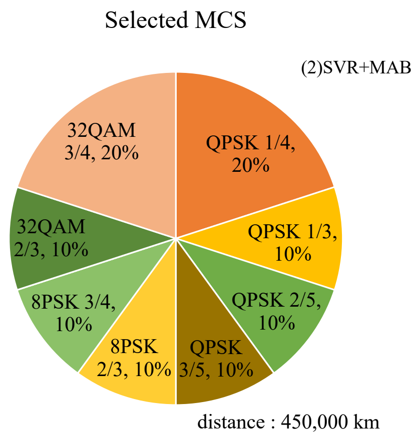


(b) Selected TCP algorithm by SVR+PSO.

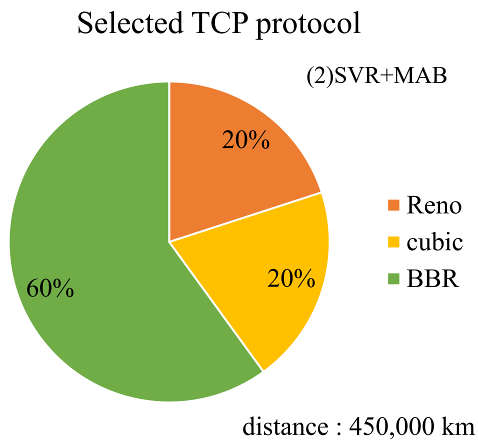


(c) Selected MTU by SVR+PSO.

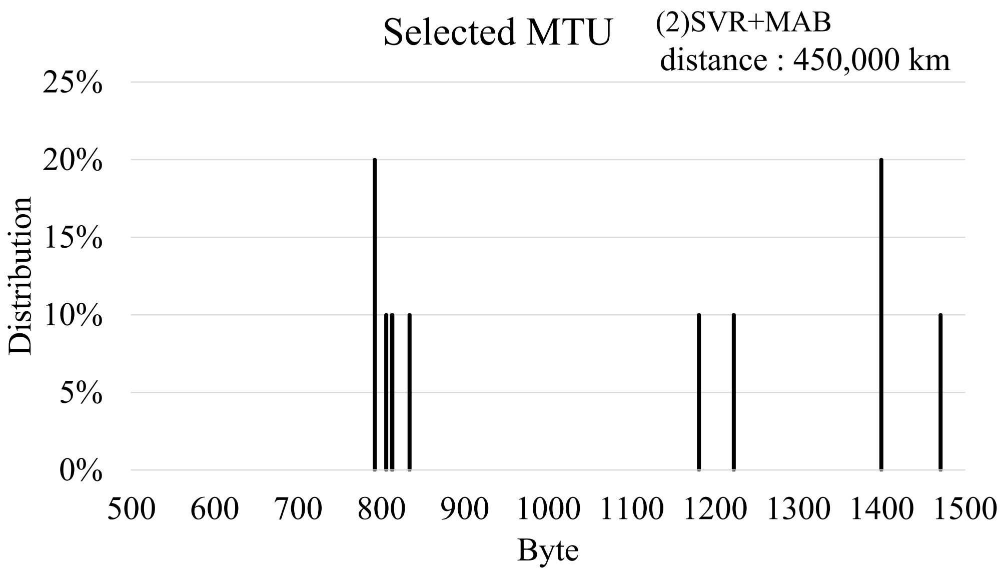
Figure 4.22: Selected parameters through iterations with algorithm (1) in 450,000 km.



(a) Selected MCS by SVR+PSO.

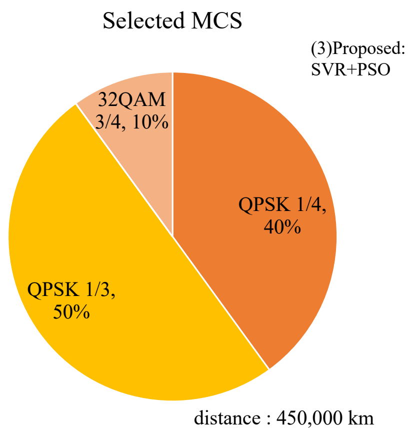


(b) Selected TCP algorithm by SVR+PSO.

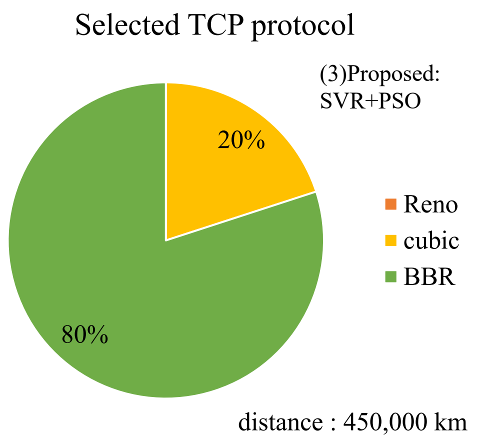


(c) Selected MTU by SVR+PSO.

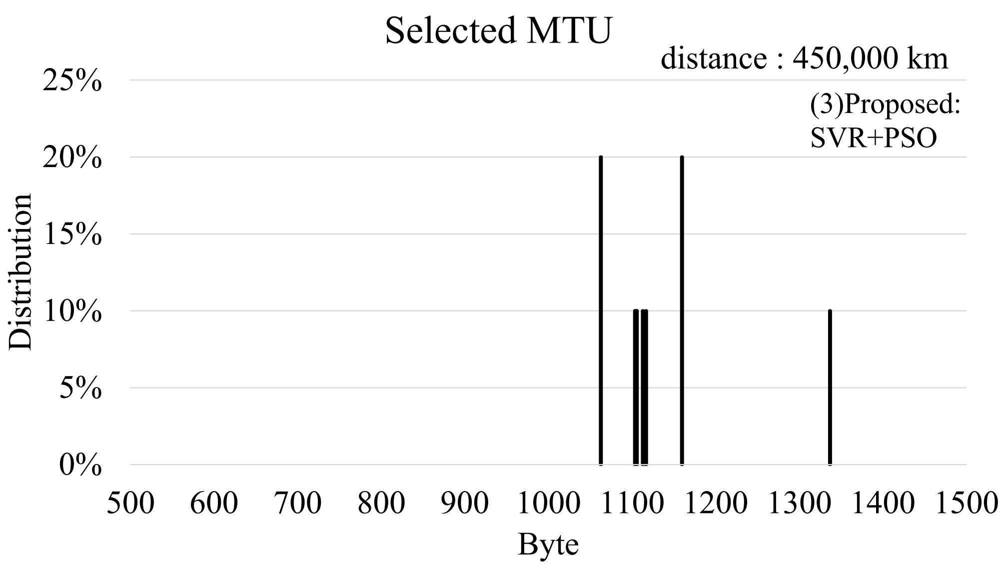
Figure 4.23: Selected parameters through iterations with algorithm (2) in 450,000 km.



(a) Selected MCS by SVR+PSO.



(b) Selected TCP algorithm by SVR+PSO.



(c) Selected MTU by SVR+PSO.

Figure 4.24: Selected parameters through iterations with algorithm (3) in 450,000 km.

Chapter 5

Simple reinforcement learning based decision

Wireless traffic and the number of wireless communication devices have increased rapidly in recent years. However, the frequency bands suitable for current technologies have already been exploited; thus, the resources are limited. Cognitive radio technologies [3, 4] have recently been developed to improve the radio resource usage of wireless networks under such situations. The basic concept of cognitive radio technology is the adaptation of the behavior of wireless systems through the recognition and learning of the radio environment. Cognitive radio systems observe and recognize the wireless network environment, make decisions, and take appropriate actions. Using this approach, various radio parameters can be optimized for the demand of wireless communications.

There are two types of cognitive radio systems; a spectrum sharing cognitive radio and a heterogeneous cognitive radio. As for spectrum-sharing cognitive radio, cognitive users utilize vacant frequency bands to improve radio resource usage. An effective channel sensing scheme is necessary for cognitive users, since the vacant radio resources changes dynamically due to other users' resource usage. Lai et al. modeled a cognitive radio as a multi-armed bandit (MAB) problem [7, 8]. The multi-armed bandit problem maximizes the total rewards provided from slot machines by optimizing the selection of slot machines that probabilistically provide rewards. In their model, channel selection of a cognitive radio is defined as a MAB problem under the assumption of probabilistic vacancy of each channel. In previous work [9], a novel MAB algorithm called tug-of-war and its application for the selection of the channel in wireless LANs have been proposed. It gave an efficient dynamic spectrum-sharing for cognitive radio.

On the other hand, recent wireless devices are equipped with multiple wireless interfaces like smartphones. Various wireless networks such as 3G, LTE and Wi-Fi are available. Moreover, like access points in Wi-Fi, a mobile terminal can choose from multiple access networks. Ideally, in such a heterogeneous network, a user may choose the best wireless network through gathering various information on each network. Several literatures have studied heterogeneous wireless network selection. There are two types of approaches: network-centric and user-centric. In the network centric approach [10–12], the decision of the selection is done at a central controller. It assumes the detail information of each network status is available at the central controller. However, it is difficult to exchange information among various wireless networks which are operated independently. Therefore, in this section, we focus on the selection of wireless networks at the mobile terminal.

Several studies investigated the wireless network selection at the mobile terminal [13–16]. Most of them needs various information of networks, or computational capacity. Authors have studied the optimizations of wireless systems based on cognitive cycle using machine learning [17]. In [17], mobile terminals gather various information of radio environments and communication performance, builds the performance model by machine learning. However, it is not always possible for mobile devices to gather information of all networks, nor to spear battery resources for complex calculations. It is important for the heterogeneous network selection to seek the better solution as much as and as fast as possible under the limited information about the networks. At the same time, it is also important for mobile devices to suppress the complexity of calculations for making decisions. These constraint and requirement are closely similar to those in the MAB problem. In the MAB problem, a player of slot machines tries to obtain maximum reward through finite trials. Therefore, the MAB problem approach can be the aid of heterogeneous network selection.

In this section, we propose an efficient wireless network selection technique of cognitive radio using the tug-of-war MAB algorithm. After the formulation of the algorithm is shown, an implementation and experimental results using a wireless device with Wi-Fi and LTE are shown. Then, as another example of the application of the proposed algorithm, channel selection in IoT devices is implemented and verified through computer simulations.

5.1 Wireless System Optimization as MAB problem

The multi-armed bandit (MAB) problem [19] is a simple machine learning problem, where a player attempts to obtain the maximum reward from multiple slot machines. The aim of the MAB is to decide which slot machine should be selected in order to obtain maximum reward through finite trials. The assumption is that the player does not have any prior information on each slot machine. The player starts to gather information on each slot machine through trying as many slot machines as possible. Then the player estimates which slot machine has the highest expected reward, and select that slot machine to play. Through this process, the player gets more rewards. There is a trade-off between exploitation and exploration. If the player takes long time for estimation, the player can estimate the reward more precisely, though the time for play the selected slot machine becomes short. If the player takes only short time for estimation, the player can take long time to play the selected slot machine, though the reward of that slot machine might be low. Fig.5.1 shows the concept of MAB problem optimal decisioning in wireless communication systems.

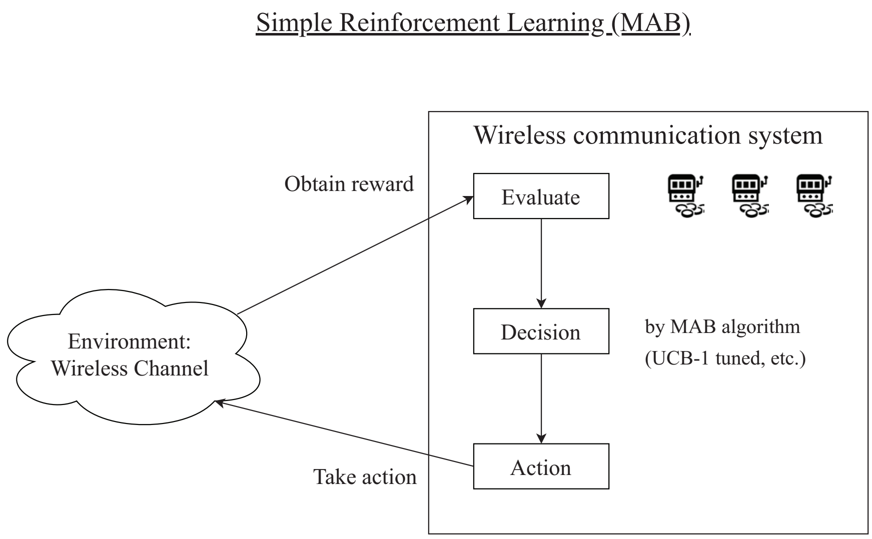


Figure 5.1: The concept of MAB problem optimal decisioning.

Figure 5.2: Network selection in cognitive radio as a multi-armed bandit problem.

In this paper, we apply the MAB problem to the wireless network selection in a cognitive mobile terminal. Fig.5.2 shows the model. The cognitive mobile terminal, as a player in the MAB problem, has N available wireless networks. Wireless networks correspond to the slot machines in the MAB problem. The terminal maintains an estimator Q_i of each wireless network i ($i = [1, N]$) according to the obtained rewards of the wireless network. The rewards can be anything such as throughput, delay, or other observables of the network performance, though the overheads to obtain them is preferable to be small. When the terminal communicates, it selects the network which can be expected highest rewards based on the estimator of each wireless network. To solve the MAB problem, we use a novel algorithm called tug-of-war (TOW) model described in the next section.

5.2 Multi-armed bandit algorithm

To solve MAB problems, several algorithms have been proposed [20–22], such as ϵ-greedy algorithm, softmax algorithm, and UCB1-tuned algorithm. Although the UCB1-tuned algorithm is known as the best algorithm among parameter-free algorithms, the tug-of-war (TOW) model [23–25] has approximately the same performance as the UCB1-tuned algorithm. The TOW model can adapt the variable environment where the reward probability changes dynamically. Therefore, the TOW model is fitted for solving the problem of decision-making in cognitive terminals.

5.3 Tug-of-war model

The tug-of-war (TOW) model is a multi-armed bandit algorithm inspired by the behavior of the amoeboid organism [23–25]. Unlike other algorithms for estimating the reward probability of each slot machine, the TOW dynamics uses a unique learning method which is equivalent to updating all machines' estimates simultaneously based on the volume conservation law. In the TOW model, the decision is made according to the displacements of the imaginary volume-conserving objects, which increase or decrease along with rewards, as shown in Fig.5.3. The TOW model of two machines are formulated as below. Imagine that the player plays a slot machine A or B at a time. When playing the machine A, if the player gets rewards, then 1 is added to an estimator Q_A, otherwise, ω is decreased (punishment). After playing time step t, the displacement of machine A, X_A $(= -X_B)$, is expressed as followings:

$$
\begin{aligned}
X_A(t+1) &= Q_A(t) - Q_B(t) + \delta(t), \qquad (5.1)\\
Q_A(t) &= N_A(t) - (1+\omega)L_A(t), \qquad (5.2)
\end{aligned}
$$

where $\delta(t)$ is a fluctuation, $N_i(t)\,(i \in \{A, B\})$ counts the number of times that machine i has been played until time t, $L_i(t)$ counts the number of punishments when playing machine i until time t. ω is a weighting parameter to be described below.

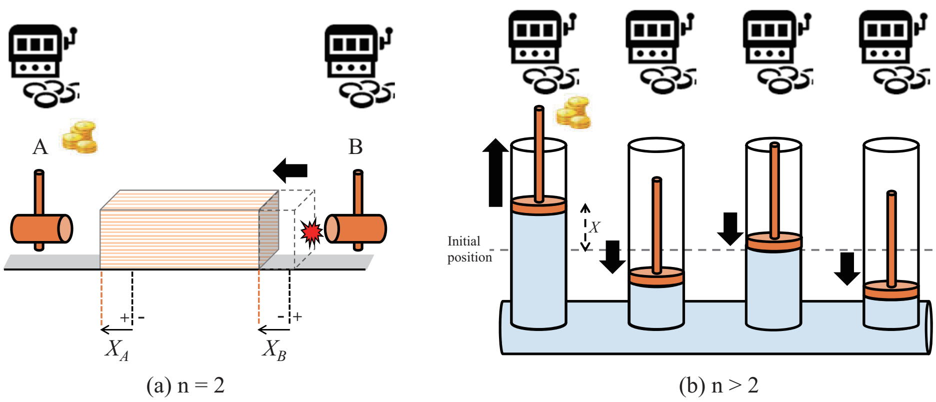


Figure 5.3: (a) The TOW model with two slot machines. A solid bar keeps its shape constant. (b) The TOW model with several slot machines. A branched cylinder is filled with uncompressive fluid.

Let probabilities of providing rewards in the machines i be $P_i\,(i \in A, B)$. Considering the ideal situation where the sum of the reward probabilities $\gamma = P_A + P_B$ is known to the player, the expected reward $Q'_i\,(i \in A, B)$ is given as

$$
\begin{aligned}
Q'_A &= N_A \frac{N_A - L_A}{N_A} + N_B \left(\gamma - \frac{N_B - L_B}{N_B}\right) \\
&= N_A - L_A + (\gamma - 1) N_B + L_B, \qquad (5.3) \\
Q'_B &= N_A \left(\gamma - \frac{N_A - L_A}{N_A}\right) + N_B \frac{N_B - L_B}{N_B} \\
&= N_B - L_B + (\gamma - 1) N_A + L_A. \qquad (5.4)
\end{aligned}
$$

If we define $Q''_j = Q'_j/(2-\gamma)$, we can obtain the difference of estimates in ideal situation as:

$$Q''_A - Q''_B = (N_A - N_B) - \frac{2}{2-\gamma}(L_A - L_B). \qquad (5.5)$$

On the other hand, the difference of Q_A and Q_B from Eq.(5.2) is given by

$$Q_A - Q_B = (N_A - N_B) - (1+\omega)(L_A - L_B). \qquad (5.6)$$

From above two equations, we can obtain the nearly optimal weighting parameter ω in terms of γ as:

$$\omega = \frac{\gamma}{2-\gamma}. \qquad (5.7)$$

If the number of machines is n $(n > 2)$, $\omega = \gamma/(2-\gamma)$ is given by $\gamma = P_1 + P_2$, where P_1 and P_2 are the first- and second- highest reward probabilities, respectively [24]. Then Eq.(5.1) can be expressed as:

$$X_i(t) = Q_i(t) - \frac{1}{n-1} \sum_{k=1,\neq i}^{n} Q_k(t) + \zeta_i(t), \qquad (5.8)$$

where ζ is a fluctuation for each slot machine. The player selects the machine which has the highest $X_i(t)$. We use the following $\zeta_i(t)$ in this paper:

$$\zeta_i(t) = A \cos\left(\frac{2\pi t}{n} + \frac{2(i-1)\pi}{n}\right), \qquad (5.9)$$

where A is the amplitude of the fluctuation.

5.4 Application of the MAB algorithm to wireless network selection

5.4.1 Heterogeneous network selection as a multi-armed bandit problem

The major challenges in the selection of wireless network at the mobile terminal in heterogeneous environments are below:

- Making efficient decision under the situation where few information of each network is available.
- Practical algorithm which can be implemented on resource-constraint mobile devices.

To overcome these issues, several studies investigated algorithms and performances of them. In [13], a non-cooperative game formulation and analysis were given for Wi-Fi and cellular network selection on the mobile terminal. Results of computer simulations showed that the game can converge to Nash equilibria. However, the assumption that the mobile terminal can get the information of other mobile user is not always possible. In [16], a reinforcement learning solution and simulation analysis were given for heterogeneous cellular networks. Even though the simulation results showed the convergence speed and the suppression of overheads, it requires feedback information from the networks, which is only feasible in the cellular networks. It is important for the mobile terminal in heterogeneous network to select the better network without the coordination from the networks. At the same time, it is also important for mobile devices to suppress the complexity of calculations for making decisions. These constraint and requirement are closely similar to those in the MAB problem.

We propose the wireless network selection technique based on the MAB problem. As described in the previous section, we use the TOW algorithm to solve the MAB problem. Fig.5.4 shows the concept of proposed wireless net-

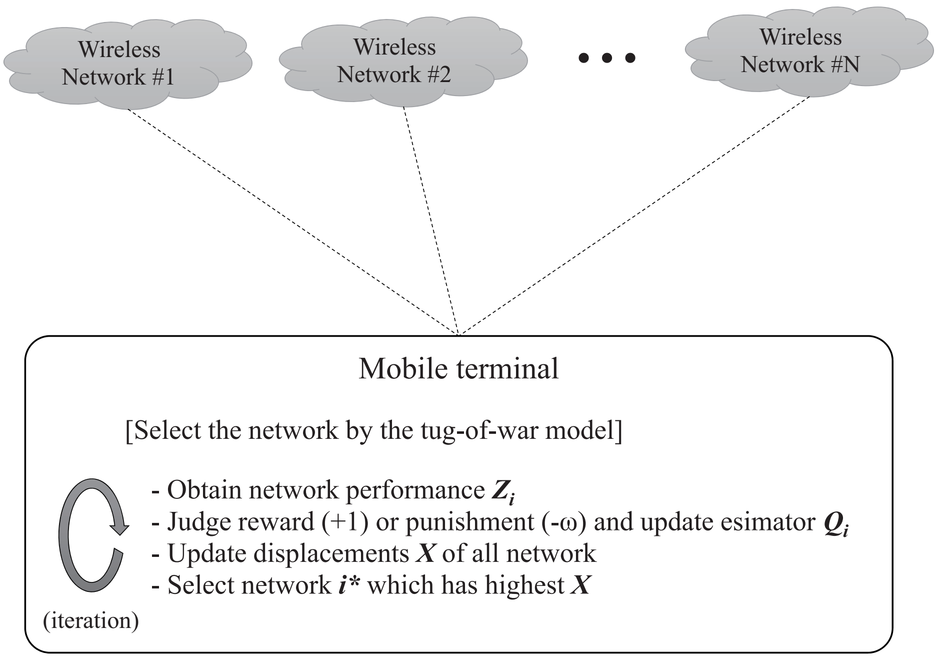


Figure 5.4: Concept of the proposed wireless system which selects the wireless network based on the MAB algorithm.

work selection. The mobile terminal, capable of connecting various kind of wireless networks S_i ($i \in \{1, 2, ...N\}$), has the function of TOW algorithm. It observes the performance of the network Z_i, where i is the selected network. Z_i can be any indexes of performance such as throughput, delay, or other metrics of wireless networks. The TOW algorithm then judges whether the reward or the punishment is to be added, by evaluating the obtained performance of network i. If the reward is given, it updates the estimator as $Q_i + 1$, otherwise, updates the estimator as $Q_i - \omega$. Then, X, the displacement of each network, is updated as described in Eq.(5.8). Note that all X_j ($j \in \{1, 2, ..., i, ..., N\}$, not only the selected network i, are updated here. The algorithm in the mobile terminal is as follows:

1. Start observing the performance of each wireless network i. All networks are monitored at least once.

2. Update Q_i and all X based on the observation of network i by Eq.(5.2) and Eq.(5.8).

3. Select a wireless network i^* with highest X_i.

4. Observe the performance of the selected network and decide whether the reward or the punishment is to be given.

5. Back to 2 and continue.

5.4.2 Implementation and Evaluation

Implementation of the proposed scheme

In order to validate the proposed method in heterogeneous wireless network environment, we implement the proposed algorithm to a wireless device and perform experiments. The TOW algorithm is installed on Ubuntu Linux in Laptop PC as a cognitive mobile terminal, which is equipped both 802.11n/ac (2.4 GHz and 5 GHz) and LTE communication module. To judge the reward or punishment (1 or $-\omega$) in Eq.(5.2), we use average throughput of wireless networks as a threshold. If the observed throughput of wireless network i is larger than the average, then the reward is given for Q_i, otherwise the punishment is given. We use the first- and second- highest reward probabilities among the networks as γ in Eq.(5.7). The experiments are conducted in and around the university building. The mobile terminal selects the wireless network from two Wi-Fi networks (2.4 GHz and 5 GHz) of the university infrastructure and LTE, to communicate to the server.

Experimental setup

We use iperf command to observe the throughput. The parameter A of the fluctuation in Eq.(5.9) is set to 5. Each iteration cycle has about 2 seconds. The locations of the experiments are (a) Laboratory room, (b) Inside the building, and (c) Outside. The average receive signal strength indicator (RSSI) of Wi-Fi is shown in Table 5.1.

Table 5.1: Average RSSI of Wi-Fi

Location	Wi-Fi 2.4 GHz	Wi-Fi 5 GHz
(a) Laboratory room	-36.0	-37.0
(b) Inside the building	-39.0	-76.0
(c) Outside	-73.0	-81.0

Verification of the proposed scheme

Fig.5.5 shows an example of network selections of the proposed algorithm. The values X of the TOW algorithm in Eq.(5.8 are also shown. In each case, after the initial trial of all wireless networks, the selection of the wireless network is converged to the highest performance network. Note that the values X of unselected networks are also updated (decreased) through iteration. Even though the estimators Q of unselected networks in Eq.(5.2) are not updated, the displacements X of all networks are updated in Eq.(5.8).

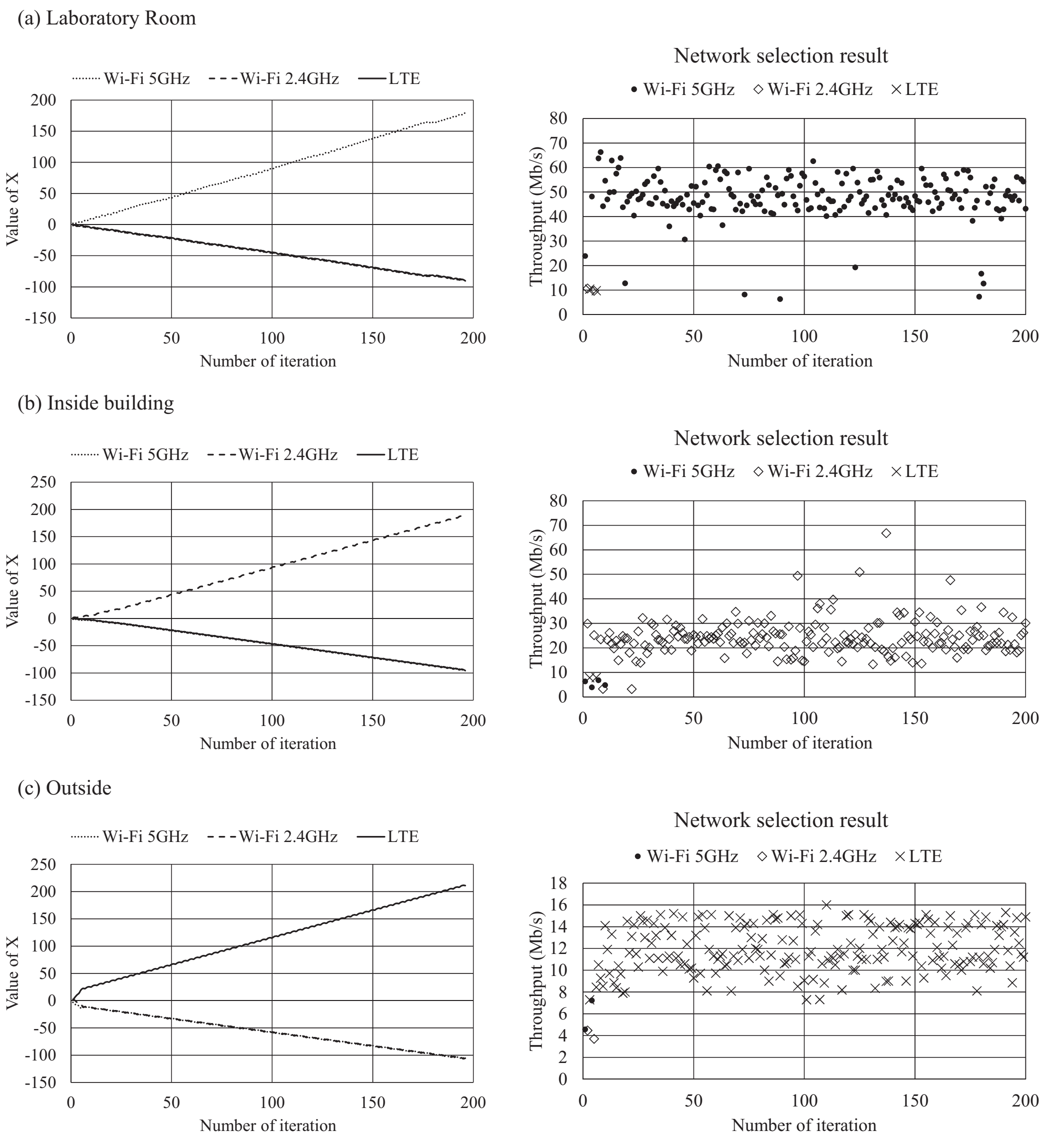


Figure 5.5: Examples of the value X in the TOW algorithm (left) and the results of network selection (right).

This is the unique characteristic of the TOW model, as described in the previous section. In the case (a) and (b), the values X of Wi-Fi 5 GHz and 2.4 GHz, which have higher RSSI and throughput in laboratory room and inside the building, become larger according to the number of iteration, while X of unselected Wi-Fi and LTE become smaller. As a result, the selection is converged to Wi-Fi 5 GHz and 2.4 GHz. In the case (c), where the signal strength from Wi-Fi access point becomes much lower, the value X of LTE becomes larger according to the number of iteration, while X of unselected Wi-Fi and LTE become smaller. As a result, the selection of the networks is converged to LTE.

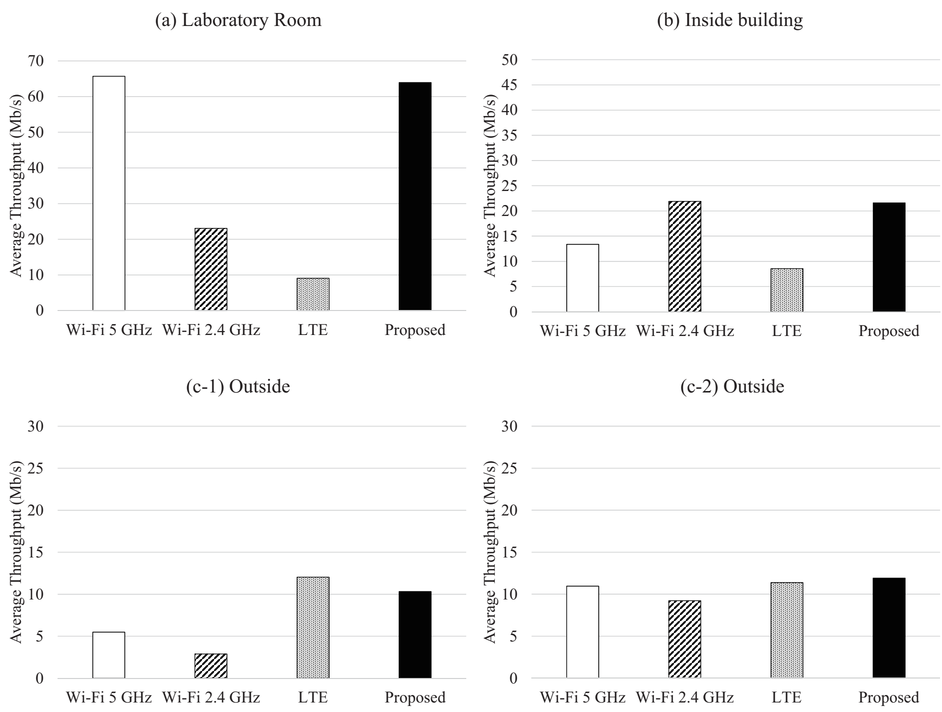


Figure 5.6: Average throughputs of each wireless networks and the proposed TOW algorithm in different environments.

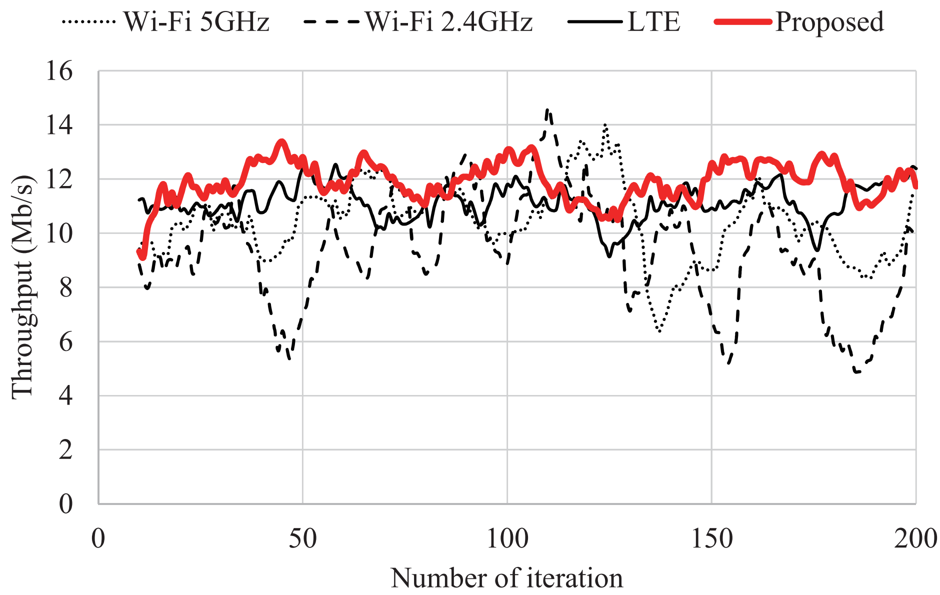


Figure 5.7: An example of throughput variation by time in (c-2).

Evaluation of the throughput performance of the proposed scheme

In order to verify the performance in heterogeneous environments, we examine the throughput in each place. Fig.5.6 shows the average throughputs of each wireless network and the proposed TOW algorithm are shown. Each experiment is continued three times, and the throughputs shown here are the average of them. The locations (c-1) and (c-2) are both outside the building but different in Wi-Fi traffic situation: (c-1) is more crowded. It is shown that the proposed TOW algorithm achieve as high average throughput among other wireless networks. In the laboratory room (a) and inside the building (b), the throughputs of the proposed system are as high as those of Wi-Fi in 5 GHz and 2.4 GHz, respectively. On the other hand, outside the building (c-1) and (c-2), where the signal strength from Wi-Fi access point becomes much lower, the throughput of the proposed system is as high as that of LTE. Moreover, as shown in the case (c-2), the proposed algorithm can achieve as high performance as LTE on average, where the differences in performance among all networks are rather small. An example of throughput variation by time is shown if Fig.5.7. Throughputs when using only Wi-Fi 2.4 GHz, 5 GHz, or LTE are shown. The values are moving averages of 10 samples. This figure shows that sometimes the throughputs of Wi-Fi are higher than LTE, though the averages are lower than that of LTE. The result shows that the proposed algorithm can estimate the probabilities of rewards among the wireless networks properly.

5.5 Application of the MAB algorithm to dynamic channel selection in IoT devices

In this section, another example of application of simple reinforcement learning based optimal decision as introduced in previous section, namely MAB problem formulation and applying TOW algorithm, is shown. The major challenges in the selection of wireless channel autonomously at the mobile terminal in massive IoT use cases are below:

- Making efficient decision under the situation where few or no information of each mobile nodes is available.

- Practical algorithm which can be implemented on resource-constraint mobile devices like IoT.

To overcome these issues, several studies investigated algorithms and performances of them. In [13], a non-cooperative game formulation and analysis were given for Wi-Fi and cellular network selection on the mobile terminal. Results of computer simulations showed that the game can converge to Nash equilibria. However, the assumption that the mobile terminal can get the information of other mobile user is not always possible. In [16], a reinforcement learning solution and simulation analysis were given for heterogeneous cellular networks. Even though the simulation results showed the convergence speed and the suppression of overheads, it requires feedback information from the networks, which is only feasible in the cellular networks. It is important for the mobile terminal in heterogeneous network to select the better network without the coordination from the networks. At the same time, it is also important for mobile devices to suppress the complexity of calculations for making decisions. These constraint and requirement are closely similar to those in the MAB problem.

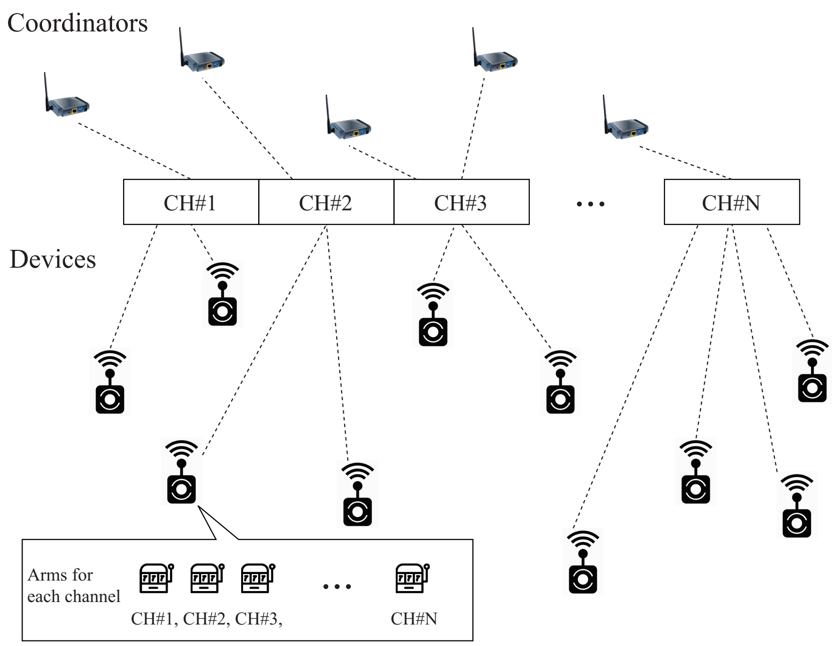


Figure 5.8: Concept of the proposed wireless system which selects the wireless network based on the MAB algorithm.

5.5.1 System model

We propose the wireless network selection technique based on the MAB problem. As described in the previous section, we use the TOW algorithm to solve the MAB problem. Fig.5.8 shows the concept of proposed wireless network selection. The mobile terminal, capable of connecting various kind of wireless networks S_i ($i \in \{1, 2, ...N\}$), has the function of TOW algorithm. It observes the performance of the network Z_i, where i is the selected network. Z_i can be any indexes of performance such as throughput, delay, or other metrics of wireless networks. The TOW algorithm then judges whether the reward or the punishment is to be added, by evaluating the obtained performance of network i. If the reward is given, it updates the estimator as $Q_i + 1$, otherwise, updates the estimator as $Q_i - \omega$. Then, X, the displacement of each network, is updated as described in Eq.(5.8). Note that all X_j ($j \in \{1, 2, ..., i, ..., N\}$, not only the selected network i, are updated here. The algorithm in the mobile terminal is as follows:

1. Start observing the performance of each wireless network i. All networks are monitored at least once.

2. Update Q_i and all X based on the observation of network i by Eq.(5.2) and Eq.(5.8).

3. Select a wireless network i^* with highest X_i.

4. Observe the performance of the selected network and decide whether the reward or the punishment is to be given.

5. Back to 2 and continue.

The algorithm used to applied is the same one, as introduced in previous section in this paper.

5.5.2 Implementation and Evaluation

In this example, the computer simulation experiments are conducted, because the massive IoT scenario requires amount of mobile devices: at least around some tens of devices which needs some scales of implementation and experiment in terms of human resources, equipments, and time.

Implementation of the proposed scheme

In order to validate the proposed method in heterogeneous wireless network environment, we implement the proposed algorithm to a wireless device and perform experiments. The TOW algorithm is installed on Ubuntu Linux in Laptop PC as a cognitive mobile terminal, which is equipped both 802.11n/ac (2.4 GHz and 5 GHz) and LTE communication module. To judge the reward or punishment (1 or $-\omega$) in Eq.(5.2), we use average throughput of wireless networks as a threshold. If the observed throughput of wireless network i is larger than the average, then the reward is given for Q_i, otherwise the punishment is given. We use the first- and second- highest reward probabilities among the networks as γ in Eq.(5.7). The experiments are conducted in and around the university building. The mobile terminal selects the wireless network from two Wi-Fi networks (2.4 GHz and 5 GHz) of the university infrastructure and LTE, to communicate to the server.

Experimental setup

In order to verify the proposed simple reinforcement algorithm in massive IoT scenario, network simulation using ns-3 [294] is conducted. The settings of simulation are shown in Table 5.2.

Two simulation setups are considered as described in 5.3.

Verification of the proposed scheme

The performance index of this simulation is set as frame success rate (FSR), the ratio of the number of received packets to the total number of transmitted

Table 5.2: Simulation settings

Parameter	Value
Area size	100 m x 100 m
Simulation duration	90 s
Simulation iterations	10 times
Number of repetition	10 times with different random seeds
Pathloss model	Free space decay
Propagation channel model	Additive white gaussian noise (AWGN)
Wireless standard	IEEE 802.15.4
Radio frequency	2.4 GHz
Modulation scheme	O-QPSK
Transmission power	1 mW
Radio communication range	around 100 m
Mac protocol	CSMA/CA
Number of channels	4, 6, 8, 10
Traffic	UDP of 100 Byte/0.2 s
Number of coordinators	10
Number of devices	40
Node placement	normalized random distribution
Mobility	All nodes are stationary

packets. Fig.5.9a shows the FSR of nodes in the scenario A): coordinators are operated in fixed channels. FSR when devices select their operational channels by TOW algorithm outperforms those by ϵ-Greedy algorithm or by UCB-1 tuned algorithm. It indicates that TOW algorithm can select

Table 5.3: Two scenarios of channel selection by simple reinforcement learning based optimal decision in massive IoT.

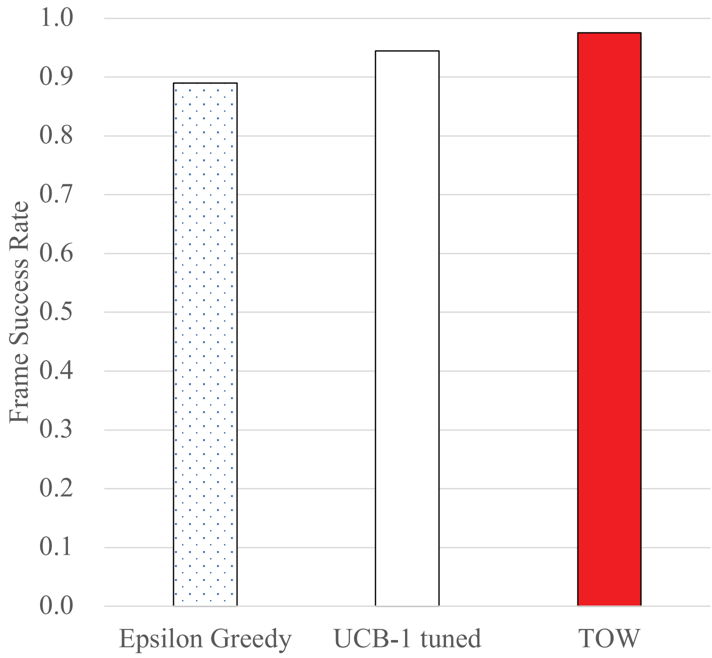
Case	Channel on coordinator	Channel on device
A	Fixed manually	Selected by MAB algorithms
B	Selected by MAB algorithms	Selected by MAB algorithms

operational channels efficiently among other MAB algorithms.

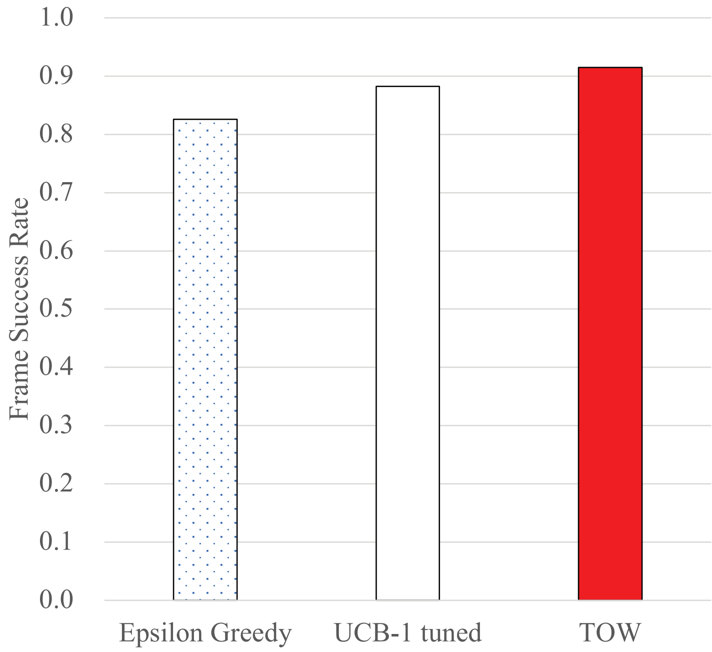
The scenario B, where not only devices but also coordinators select their operational channels autonomously, is more difficult where each nodes select optimal channels, because both coordinators and devices select operational channels independently, which gives the problem of matching in channel selection among devices and coordinators. This scenario is an example of highly distributed decision network. The result of this scenario is shown in Fig.5.9b. FSR when devices select their operational channels by TOW algorithm outperforms those by ϵ-Greedy algorithm or by UCB-1 tuned algorithm. It indicates that TOW algorithm can select operational channels efficiently among other MAB algorithms, even in more distributed selection scheme than in the scenario B.

Fig.5.10–5.12 show examples of channel selection results at all devices in each algorithm in the scenario A (channels of coordinators are fixed). From these figure,

- The selected channels show convergence in the algorithm of TOW.
- The selected channels in the algorithm of UCB-1 also show convergence but a little fluctuated over time, compared to those in TOW, which leads the overhead of channel switching.
- The selected channels in the algorithm of ϵ-Greedy are unstable and fluctuated compared to the rest of two algorithms. It comes from the "greedy" exploration characteristic of the algorithm.



(a) Frame success rate of each algorithm in scenario A.



(b) Frame success rate of each algorithm in scenario B.

Figure 5.9: Average frame success rate of wireless nodes.

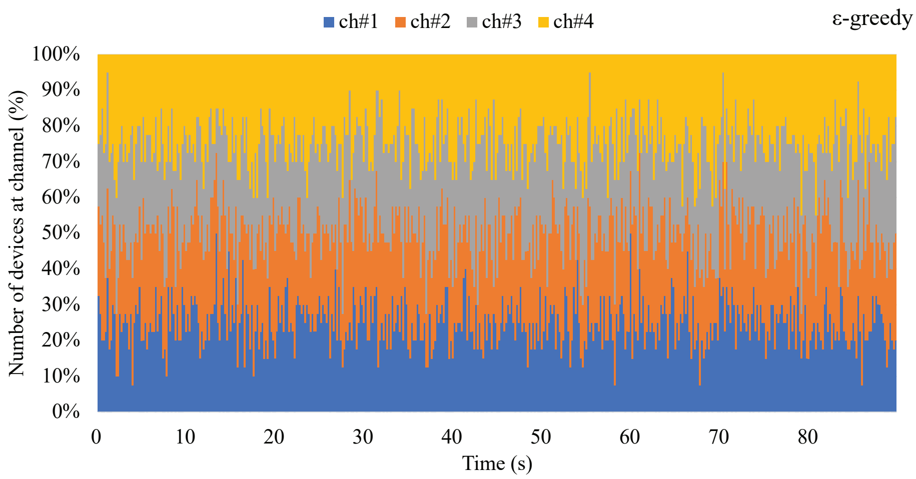


Figure 5.10: The channel selection results at all devices in the scenario A with ϵ-Greedy algorithm.

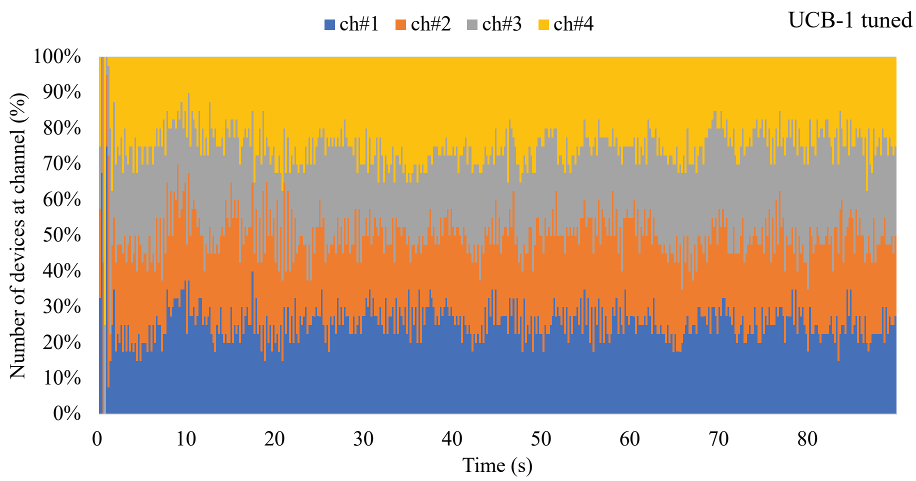


Figure 5.11: The channel selection results at all devices in the scenario A with UCB-1 tuned algorithm.

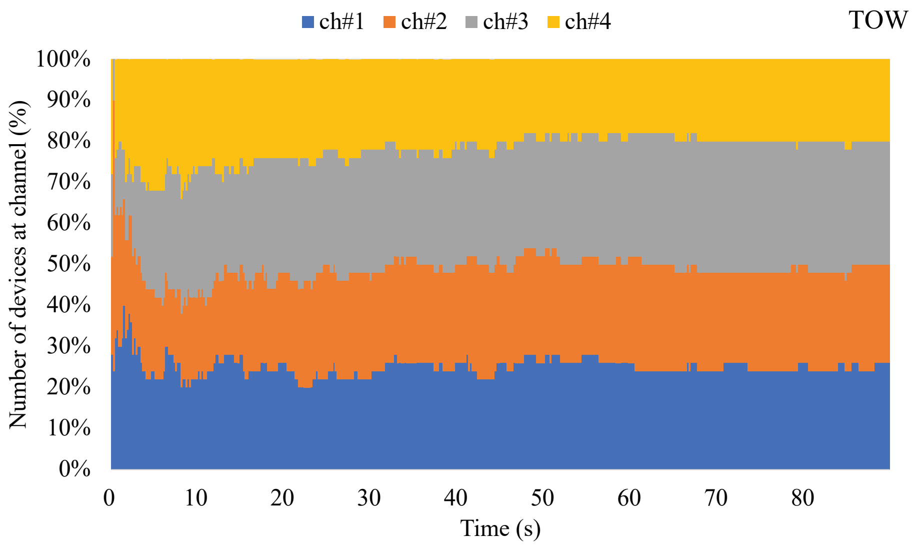


Figure 5.12: The channel selection results at all devices in the scenario A with TOW algorithm.

Fig.5.13–5.15 show examples of channel selection results at all devices in each algorithm in the scenario B (channels of coordinators are also selected by the same algorithm as devices). Note that in this scenario B, compared to the scenario A, the channel selection becomes more difficult. It is because the channel selection is conducted not only devices but also coordinators, which makes this wireless communication system more complex. There is a problem of channel matching among devices and coordinators, which may lead some oscillations among channels, resulting in poor communication performance. From these figure,

- On average, FSR in each algorithm is lower than that in the scenario A, which indicates that the channel selection in this scenario is more difficult than in the scenario A, as indicated above.
- The selected channels show convergence in the algorithm of TOW.
- The selected channels in the algorithm of UCB-1 also show convergence to equality but a little fluctuated over time, compared to those in TOW,

which leads the overhead of channel switching.

- The selected channels in the algorithm of ϵ-Greedy are unstable and fluctuated compared to the rest of two algorithms. It comes from the "greedy" exploration characteristic of the algorithm.

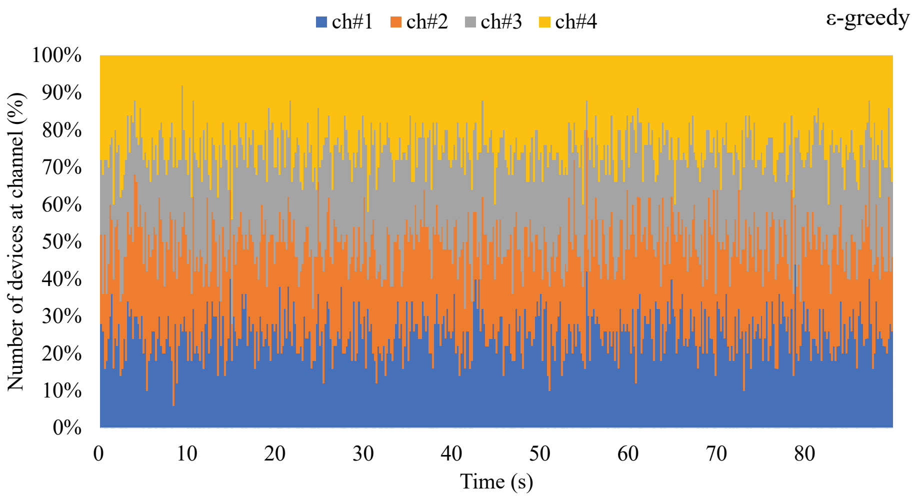


Figure 5.13: The channel selection results at all devices in the scenario B with ϵ-Greedy algorithm.

Fig.5.16 shows examples of channel selection results at all devices in each algorithm in the scenario A (channels of coordinators are fixed).

Fig.5.17 shows examples of channel selection results at all devices in each algorithm in the scenario B (channels of coordinators are also selected by the same algorithm as devices).

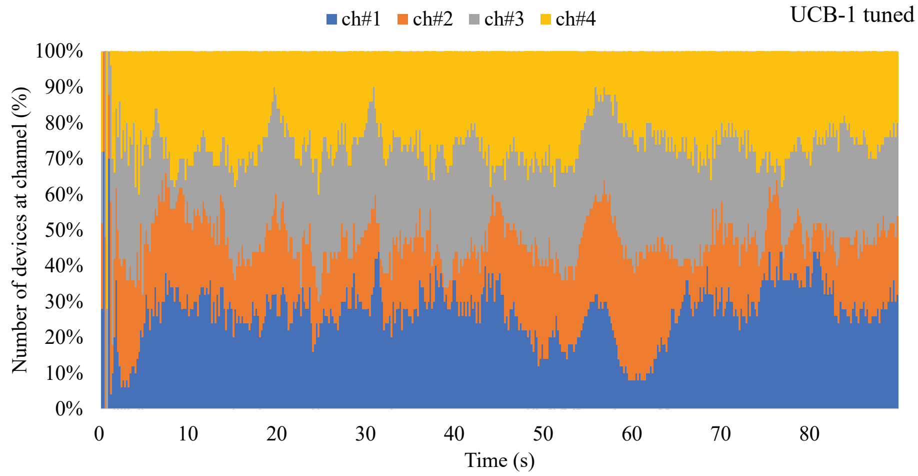


Figure 5.14: The channel selection results at all devices in the scenario B with UCB-1 tuned algorithm.

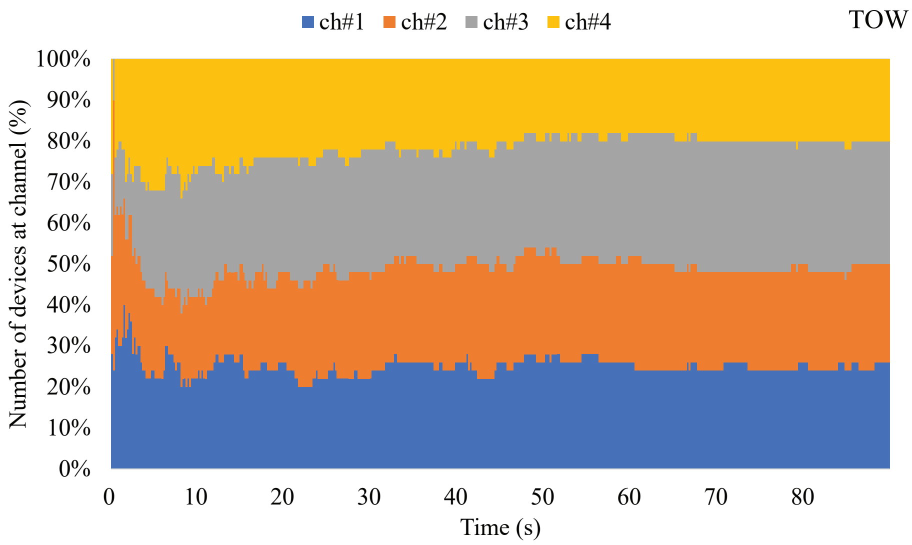
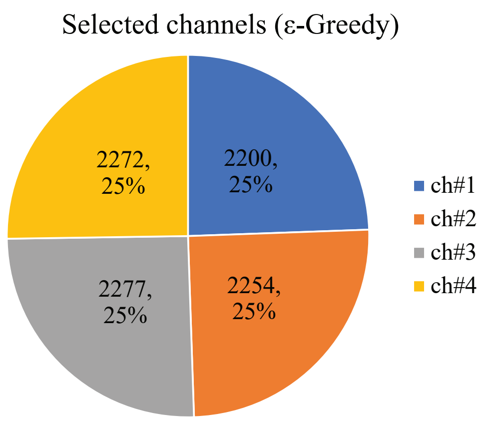
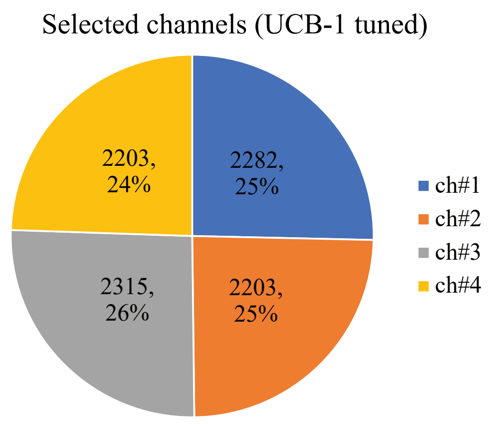


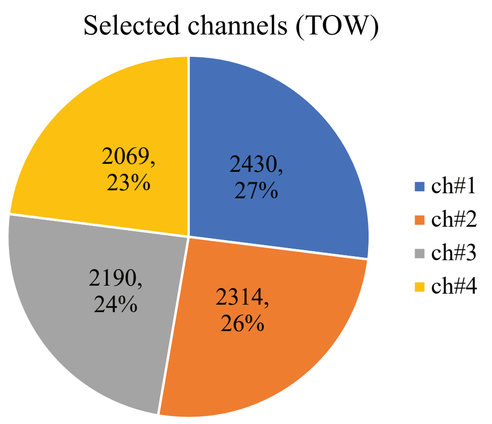
Figure 5.15: The channel selection results at all devices in the scenario B with TOW algorithm.



(a) The channel selection results at all devices in the scenario A with ϵ-Greedy algorithm.

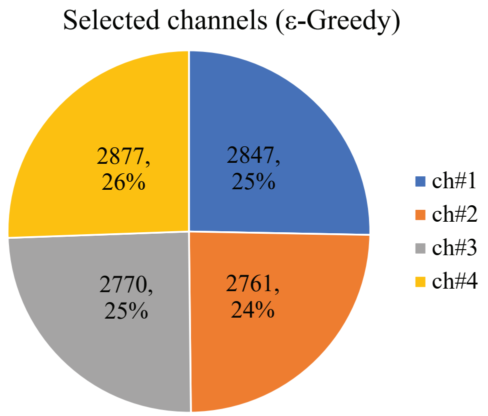


(b) The channel selection results at all devices in the scenario A with UCB-1 tuned algorithm.

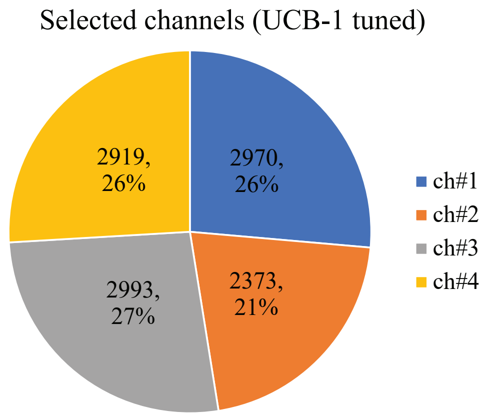


(c) The channel selection results at all devices in the scenario A with TOW algorithm.

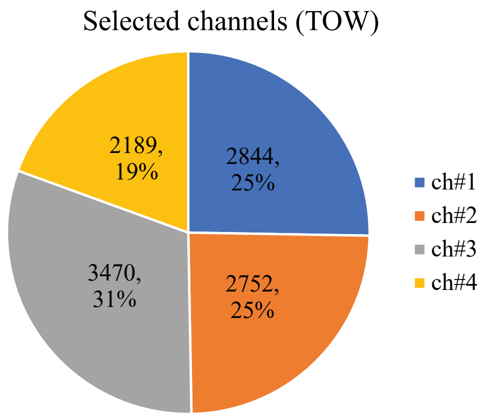
Figure 5.16: Channel selection at STAs in scenario A.



(a) The channel selection results at all devices in the scenario B with ϵ-Greedy algorithm.



(b) The channel selection results at all devices in the scenario B with UCB-1 tuned algorithm.



(c) The channel selection results at all devices in the scenario B with TOW algorithm.

Figure 5.17: Channel selection at STAs in scenario B.

5.6 Summary

Due to the advance in mobile wireless systems and the scarcity of the frequency spectrum, it is necessary for mobile devices to utilize wireless networks in a heterogeneous environment. Wireless network selection is one of the realistic problem in current mobile devices. Considering the limitation of hardware and software complexity in mobile deices, an efficient wireless network selection in a cognitive way is important. The multi-armed bandit problem is a simple machine learning problem and applicable for cognitive radio problem. In this section, we proposed a simple but powerful wireless network selection technique using the novel multi-armed bandit algorithm called tug-of-war. Through the implementation and experiments, the effectiveness of the proposed algorithm in a real heterogeneous wireless environment was shown. Since the proposed technique is based on a simple and light-weight algorithm, it can be applied to not only smartphones, but also IoT devices where the system resources are more limited. It is important to investigate the performance in real application services, and the candidates of reward indexes.

Chapter 6

Conclusion

6.1 Summary of this paper

This paper consists of six sections: introduction of this paper is in Chapter 1, the issues of current and future wireless communication systems are pointed out in Chapter 2, then the introducing machine learning to overcome these issues are discussed in Chapter 3. The proposed two schemes, supervised learning based modeling and optimization based decisioning, and simple reinforcement learning based decisioning are introduced in Chapter 4 and 5 respectively, where the evaluation and the verification of the proposed schemes through real device experiments and computer simulations are also shown in these sections. Following are the brief summary of Chapters in this paper.

In Chapter 1, the background and the aim of this research are shown. Advancement of wireless communication technologies have brought us enormous positive change all over the world. Yet, from the viewpoint of exploiting its capability, there are still some issues to maximize their performance, which rises two simple questions. One question is that how to build models of today's and future complex wireless systems. Another question is that how to decide optimal action by using the models of wireless communication systems.

In Chapter 2, issues of today's and future wireless communication systems are indicated, especially in terms of optimizing their performance. Classical mathematical formulation based optimization scheme cannot be applied any more for today's complex wireless communication systems, because the complexity of the systems prevent building mathematical model. It opens the window of applications of machine learning technologies to optimize performances of wireless communication systems.

In Chapter 3, application of machine learning technologies and its issues are discussed, then the proposed schemes in this research are introduced. Classical optimization of performances of wireless communication systems is based on mathematical formulation, which cannot be applied to today's complex, time-varying system usage. A data-driven modeling, by machine learning, is an aid for this issue. Deep reinforcement learning, deep learning based modeling and reinforcement learning based decision for action, is a state-of-the-art scheme in recent research fields. It has recently applied in the field of wireless communication a lot. However, it does not mean that all issues of current and future wireless systems can be solved by it. There still exist future works to be pointed out. One point is to seek alternatives for modeling to realize continuous function based modeling to obtain better solutions for continuous systems. Another point is to seek the feasible yet effective scheme if the amount of available information is small, like in IoT systems. Corresponding to these points, two novel schemes are proposed which are different not only from classical mathematical optimization but also from current state-of-the-art deep reinforcement learning approach.

In Chapter 4, one of the proposed schemes, supervised learning based modeling and optimization, is formulated and examined through experiments. It uses some amount of information to build the model of the wireless communication system, and obtain optimal parameters by an optimization algorithm. This is based on cognitive cycle using machine learning. It uses by supervised machine learning algorithm to build the performance model of the systems, obtains the optimal parameters by solving the optimization problem, takes action according to the decision, and updates the performance model in online manner. Two applications are shown: IEEE 802.11 WLAN and space communication. Through both real-world experiments and computer simulations, the validity of the proposed scheme is confirmed.

In Chapter 5, another proposed scheme, simple yet easily-implementable reinforcement learning, by MAB problem formulation is formulated and examined through experiments. By using a novel, light-weight, and distributed TOW algorithm, it realizes adaptive learning wireless communication systems whose capabilities in software and/or hardware are limited like IoT. Two applications are shown: heterogeneous network selection and channel selection in massive IoT. Through both real-world experiments and computer simulations, the validity of the proposed scheme is confirmed.

These results show the effectiveness and feasibility of the proposed schemes. Various applications based on the proposed schemes are currently being developed [297–300], which proves that this research opened a new field of application of online machine learning technologies to optimize today's and future complex wireless communication systems.

6.2 Achievement of this paper

The major achievements of this paper are:

1. Provided two novel approaches from wide viewpoint of current application of machine learning to wireless communication.

2. Proposed scheme using supervised learning and optimization gives a better alternative of deep reinforcement learning especially when parameters are continuous.

3. Proposed scheme using simple reinforcement learning based on TOW, a ligh-weight MAB algorithm, provides feasible solution to increase performances of wireless communication systems where amount of available information is small like IoT.

6.3 Concluding remarks

There are some future works which will widen this research. As for application to space communication networks, an example verfied in this paper is quite simple one. This application will develop an optimization problem like relay network in space, multiple interfaces for radio and optical communication, etc. In terms of adaptability, rapidly time-varying environment, such as vehicle communication networks, will require an advancement of learning technologies proposed in this paper. Some researches such as introducing forgetting factor to reinforcement learning algorithm, are being conducted. Through various applications are developed, more fundamental aspects will be revealed such as an application for convex/non-convex problem, etc.

References

Bibliography

[1] Y. eCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature 521, pp. 436-–444 (2015).

[2] G. Bianchi, “Performance analysis of the IEEE 802.11 distributed coordination function,” IEEE Journal on Selected Areas in Communications, vol. 18, no. 3, pp. 535-547, March 2000.

[3] J. Mitola and G. Q. Maguire, “Cognitive radio: Making software radios more personal,” IEEE Pers. Commun., vol. 6 no. 4, pp. 13–18, Apr. 1999.

[4] S. Haykin, “Cognitive radio: Brain-empowered wireless communications,” IEEE Journal on Selected Areas in Communications, vol. 23 no. 2, pp. 201–220, Feb. 2005.

[5] J. Mitola, “Cognitive radio architecture evolution,” Proc. IEEE, vol. 97, no. 4, pp. 626–641, Apr. 2009.

[6] “AlphaGo: Mastering the ancient game of Go with Machine Learning,” `https://ai.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html`

[7] L. Lai, H.E. Jiang, and H.V. Poor, “Medium access in cognitive radio networks: A competitive multi-armed bandit frame work,” *Proc. of IEEE 42 th Asilomar Conference on Signals, System and Computer*, pp. 98–102, 2008.

[8] L. Lai, H.E. Jiang, and H.V. Poor, "Cognitive medium access: Exploration, exploitation, and competition," *IEEE Trans. on mobile computing*, vol. 10, no. 2, pp. 239–253, 2011.

[9] K. Kuroda, H. Kato, S.-J. Kim, M. Naruse, and M. Hasegawa, "Improving throughput using multi-armed bandit algorithm for wireless LANs," *NOLTA*, vol. 9, no. 1, pp. 74–81, 2018.

[10] S. Singh, J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 888–901, 2014.

[11] C. Liu, M. Li, S. V. Hanly, P. Whiting, "Joint downlink user association and interference management in two-tier HetNets with dynamic resource partitioning," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1365–1378, 2017.

[12] V. Sagar, R. Chandramouli, K. P. Subbalakshmi, "Software defined access for HetNets," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 84–89, 2016.

[13] A. Keshavarz-Haddad, E. Aryafar, M. Wang, M. Chiang, "HetNets selection by clients: Convergence efficiency and practicality," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 406–419, 2017.

[14] E. Aryafar, A. Keshavarz-Haddad, M. Wang, M. Chiang, "RAT selection games in HetNets," *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, pp. 998–1006, 2013.

[15] X. Wang, J. Li, L. Wang, C. Yang, Z. Han, "Intelligent User-Centric Network Selection: A Model-Driven Reinforcement Learning Framework," *IEEE Access*, vol. 7, pp. 21645–21661, 2019.

[16] D. D. Nguyen, H. X. Nguyen, L. B. White, "Reinforcement Learning With Network-Assisted Feedback for Heterogeneous RAT Selection," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, 2017.

[17] K. Oshima, T. Kobayashi, Y. Taenaka, K. Kuroda, M. Hasegawa, "Wireless network optimization method based on cognitive cycle using machine learning," *IEICE ComEX*, vol. 7, no. 7, pp. 278–283, 2018.

[18] The 3rd Generation Partnership Project (3GPP), `https://www.3gpp.org/specifications/specifications`

[19] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin of the American*, vol. 58, pp. 527–535, 1952.

[20] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, Cambridge, Massachusetts London, England, 1998 .

[21] D. Daw, P. O'Doherty, P. Dayan, B. Seymour, and J. Dolan, "Cortical substrates for exploratory decisions in humans," *Nature*, vol. 441, pp. 876–879, 2006.

[22] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, pp. 235–256, 2002.

[23] S.-J. Kim and M. Aono, "Amoeba-inspired algorithm for cognitive medium access," *NOLTA*, vol. 5, no. 2, pp. 198–209, 2014.

[24] S.-J. Kim, M. Aono, and E. Nameda, "Efficient decision-making by volume-conserving physical object," *New Journal of Physics*, vol. 17, 083023, 2015.

[25] S.-J. Kim, M. Aono, and E. Nameda, "Decision maker based on atomic switches," *AIMS Material Science*, vol. 3, no. 1, pp. 245–259, 2016.

[26] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. Chen, L. Hanzo, "Machine Learning Paradigms for Next-Generation Wireless Networks," IEEE Wireless Commun., vol. 24, no. 2, pp. 98–105, Apr. 2017.

[27] Z. Zhao, *et al.*, "Autonomic communications in software-driven networks," IEEE J. Sel. Areas Commun., vol. 35, no. 11, pp. 2431–2445, Nov. 2017.

[28] Radio Spectrum Policy Group, "Report on Collective Use of Spectrum and Other Sharing Approaches," RSPG11-392, Nov. 2011.

[29] CEPT Electronic Communications Committee, "ECC Report 205 – Licensed Shared Access (LSA)," ECC Report 205, Feb. 2014.

[30] Reconfigurable Radio Systems (RRS); Information elements and protocols for the interface between LSA Controller (LC) and LSA Repository (LR) for operation of Licensed Shared Access (LSA) in the 2 300 MHz - 2 400 MHz band, ETSI TS 103 379 V1.1.1, Apr. 2017.

[31] S. Haykin, P. Setoodeh, S. Feng, D. Findlay, "Cognitive Dynamic System as the Brain of Complex Networks," IEEE J. Sel. Areas Commun., vol. 34 issue 10, pp. 2791–2800, Oct. 2016.

[32] M. Hasegawa, H. Hirai, K. Nagano, H. Harada, K. Aihara, "Optimization for Centralized and Decentralized Cognitive Radio Networks," Proc. of the IEEE, vol. 102 no. 4, pp. 574–584, Apr. 2014.

[33] Y. Kon, K. Hashiguchi, M. Ito, M. Hasegawa, K. Ishizu, H. Murakami, H. Harada, "Autonomous Throughput Improvement Scheme Using Machine Learning Algorithms for Heterogeneous Wireless Networks Aggregation," IEICE Trans. Commun., vol. E95.B no. 4, pp. 1143–1151, Apr. 2012.

[34] IEEE Standard for Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks, IEEE Std.1900.4, 2009.

[35] M. Bkassiny, Y. Li, S. K. Jayaweera, "A Survey on Machine-Learning Techniques in Cognitive Radios," IEEE Commun. Surveys and Tutorials, vol. 15, issue 3, pp. 1135–1159, 2012.

[36] "Dynamic Rate and Channel Selection in Cognitive Radio Systems," IEEE J. Sel. Areas Commun., vol. 33, no. 5, pp. 910–921, May 2015.

[37] D. Shiung, Y. Yang, "Rate enhancement for cognitive radios using the relationship between transmission rate and signal-to-interference ratio statistics," IET Commun., vol 7, issue 18, Dec. 2013.

[38] J. Lehtomaki, M. Benitez, K. Umebayashi, M. Juntti, "Improved Channel Occupancy Rate Estimation," IEEE Trans. on Commun. vol. 63, issue 3, Mar. 2015.

[39] A. Goldsmith, S. Chua, "Adaptive coded modulation for fading channels," IEEE Trans. on Commun., vol. 46, no. 5, May 1998.

[40] D. Gesbert, S. Kiani, A. Gjendemsjo, G. Oien "Adaptation, Coordination, and Distributed Resource Allocation in Interference-Limited Wireless Networks," Proc. of the IEEE, vol. 95, no. 12, Dec. 2007.

[41] J. Lu, T. T. Tjhung, F. Adachi, C. L. Huang, "BER Performance of OFDM-MDPSK System in Frequency-Selective Rician Fading with Diversity Reception," IEEE Trans. on Veh. Technol., vol. 49 , no. 4, pp. 1216-1225, Jul. 2000.

[42] S. T. Chung, A. J. Goldsmith, "Degrees of Freedom in Adaptive Modulation: A Unified View," IEEE Trans. Commun., vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[43] "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," IEEE Trans. on Mobile Computing, vol. 1, pp.278–291, Dec. 2002.

[44] E. Peh, Y. Liang, Y. Guan, Y. Zeng, “Optimization of Cooperative Sensing in Cognitive Radio Networks: A Sensing-Throughput Tradeoff View,” IEEE Trans. Veh. Technol., vol. 58, no. 9, pp. 5294–5299, Nov. 2009.

[45] B.E. Boser, I.M. Guyon, V.N. Vapnik, “A training algorithm for optimal margin classifiers,” Proc. 5th Annual Computational Learning Theory, pp. 144–152, 1992.

[46] V. Vapnik, S. Golowich, A. Smola, “Support vector method for function approximation, regression estimation, and signal processing,” Advances in Neural Information Processing Systems, 1996.

[47] J. Kennedy, R. Eberhart, “Particle swarm optimization,” IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948, 1995.

[48] V. Kadirkamanathan, K. Selvarajah, P. J. Fleming, "Stability Analysis of the Particle Dynamics in Particle Swarm Optimizer,” IEEE Trans. on Evolutionary Computation, vol. 10, no. 3, Jun. 2006.

[49] K. Oshima, T. Kobayashi, Y. Taenaka, K. Kuroda, M. Hasegawa, “Wireless network optimization method based on cognitive cycle using machine learning,” IEICE Commun. Exp. vol. 7, no. 7, pp.278–283, 2018.

[50] Scalable Network Technologies, Inc., `https://www.scalable-networks.com`

[51] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, F. Kojima, “Big Data Analytics, Machine Learning, and Artificial Intelligence in Next-Generation Wireless Networks,” IEEE Access, vol. 6, issue , pp. 32328–32338, 2018

[52] M. Chen, U. Challita, W. Saad, C. Yin, M. Debbah, “Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial,”

IEEE Communications Surveys & Tutorials, vol. 21, issue 4, pp. 3039–3071, 2019

[53] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. -C. Chen, L. Hanzo, "Thirty Years of Machine Learning: The Road to Pareto-Optimal Wireless Networks," IEEE Communications Surveys & Tutorials, vol. 22, issue 3, pp. 1472–1514, 2020

[54] M. Kulin, T. Kazaz, I. Moerman, E. De Poorter, "End-to-End Learning From Spectrum Data: A Deep Learning Approach for Wireless Signal Identification in Spectrum Monitoring Applications," IEEE Access, vol. 6, issue , pp. 18484–18501, 2018

[55] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, K. Mizutani, "State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems," IEEE Communications Surveys & Tutorials, vol. 19, issue 4, pp. 2432–2455, 2017

[56] T. Park, N. Abuzainab, W. Saad, "Learning How to Communicate in the Internet of Things: Finite Resources and Heterogeneity," IEEE Access, vol. 4, issue , pp. 7063–7073, 2016

[57] M. A. Alsheikh, S. Lin, D. Niyato, H. Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications," IEEE Communications Surveys & Tutorials, vol. 16, issue 4, pp. 1996–2018, 2014

[58] Q. Mao, F. Hu, Q. Hao, "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey," IEEE Communications Surveys & Tutorials, vol. 20, issue 4, pp. 2595–2621, 2018

[59] R. C. Daniels, C. M. Caramanis, R. W. Heath, "Adaptation in Convolutionally Coded MIMO-OFDM Wireless Systems Through Supervised

Learning and SNR Ordering," IEEE Transactions on Vehicular Technology, vol. 59, issue 1, pp. 114–126, 2010

[60] J. Park, S. Samarakoon, M. Bennis, M. Debbah, "Wireless Network Intelligence at the Edge," Proceedings of the IEEE, vol. 107, issue 11, pp. 2204–2239, 2019

[61] J. Wang, X. Zhang, Q. Gao, H. Yue, H. Wang, "Device-Free Wireless Localization and Activity Recognition: A Deep Learning Approach," IEEE Transactions on Vehicular Technology, vol. 66, issue 7, pp. 6258–6267, 2017

[62] Y. Sun, M. Peng, Y. Zhou, Y. Huang, S. Mao, "Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues," IEEE Communications Surveys & Tutorials, vol. 21, issue 4, pp. 3072–3108, 2019

[63] I. Ahmad, M. Basheri, M. J. Iqbal, A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," IEEE Access, vol. 6, issue , pp. 33789–33795, 2018

[64] A. Zappone, M. Di Renzo, M. Debbah, "Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both?," IEEE Transactions on Communications, vol. 67, issue 10, pp. 7331–7376, 2019

[65] H. Ye, G. Y. Li, B. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," IEEE Wireless Communications Letters, vol. 7, issue 1, pp. 114–117, 2018

[66] J. Joung, "Machine Learning-Based Antenna Selection in Wireless Communications," IEEE Communications Letters, vol. 20, issue 11, pp. 2241–2244, 2016

[67] D. A. Tran, T. Nguyen, "Localization In Wireless Sensor Networks Based on Support Vector Machines," IEEE Transactions on Parallel and Distributed Systems, vol. 19, issue 7, pp. 981–994, 2008

[68] R. W. Ouyang, A. K. Wong, C. Lea, M. Chiang, "Indoor Location Estimation with Reduced Calibration Exploiting Unlabeled Data via Hybrid Generative/Discriminative Learning," IEEE Transactions on Mobile Computing, vol. 11, issue 11, pp. 1613–1626, 2012

[69] S. Mahfouz, F. Mourad-Chehade, P. Honeine, J. Farah, H. Snoussi, "Target Tracking Using Machine Learning and Kalman Filter in Wireless Sensor Networks," IEEE Sensors Journal, vol. 14, issue 10, pp. 3715–3725, 2014

[70] L. Liang, H. Ye, G. Y. Li, "Toward Intelligent Vehicular Networks: A Machine Learning Framework," IEEE Internet of Things Journal, vol. 6, issue 1, pp. 124–135, 2019

[71] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, K. Chan, "Adaptive Federated Learning in Resource Constrained Edge Computing Systems," IEEE Journal on Selected Areas in Communications, vol. 37, issue 6, pp. 1205–1221, 2019

[72] S. Rajasegarar, C. Leckie, J. C. Bezdek, M. Palaniswami, "Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks," IEEE Transactions on Information Forensics and Security, vol. 5, issue 3, pp. 518–533, 2010

[73] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, D. Tujkovic, "Deep Learning Coordinated Beamforming for Highly-Mobile Millimeter Wave Systems," IEEE Access, vol. 6, issue , pp. 37328–37348, 2018

[74] M. Chen, W. Saad, C. Yin, "Virtual Reality Over Wireless Networks: Quality-of-Service Model and Learning-Based Resource Management,"

IEEE Transactions on Communications, vol. 66, issue 11, pp. 5621–5635, 2018

[75] U. Challita, L. Dong, W. Saad, “Proactive Resource Management for LTE in Unlicensed Spectrum: A Deep Learning Perspective,” IEEE Transactions on Wireless Communications, vol. 17, issue 7, pp. 4674–4689, 2018

[76] C. Zhang, P. Patras, H. Haddadi, “Deep Learning in Mobile and Wireless Networking: A Survey,” IEEE Communications Surveys & Tutorials, vol. 21, issue 3, pp. 2224–2287, 2019

[77] M. Mozaffari, A. Taleb Zadeh Kasgari, W. Saad, M. Bennis, M. Debbah, “Beyond 5G With UAVs: Foundations of a 3D Wireless Cellular Network,” IEEE Transactions on Wireless Communications, vol. 18, issue 1, pp. 357–372, 2019

[78] O. Simeone, “A Very Brief Introduction to Machine Learning With Applications to Communication Systems,” IEEE Transactions on Cognitive Communications and Networking, vol. 4, issue 4, pp. 648–664, 2018

[79] F. Tang, Y. Kawamoto, N. Kato, J. Liu, “Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches,” Proceedings of the IEEE, vol. 108, issue 2, pp. 292–307, 2020

[80] X. Lu, H. Zou, H. Zhou, L. Xie, G. Huang, “Robust Extreme Learning Machine With its Application to Indoor Positioning,” IEEE Transactions on Cybernetics, vol. 46, issue 1, pp. 194–205, 2016

[81] Z. Xiao, H. Wen, A. Markham, N. Trigoni, P. Blunsom, J. Frolik, “Non-Line-of-Sight Identification and Mitigation Using Received Signal Strength,” IEEE Transactions on Wireless Communications, vol. 14, issue 3, pp. 1689–1702, 2015

[82] X. Geng, "Label Distribution Learning," IEEE Transactions on Knowledge and Data Engineering, vol. 28, issue 7, pp. 1734–1748, 2016

[83] S. Wang, H. Liu, P. H. Gomes, B. Krishnamachari, "Deep Reinforcement Learning for Dynamic Multichannel Access in Wireless Networks," IEEE Transactions on Cognitive Communications and Networking, vol. 4, issue 2, pp. 257–265, 2018

[84] H. Zou, B. Huang, X. Lu, H. Jiang, L. Xie, "A Robust Indoor Positioning System Based on the Procrustes Analysis and Weighted Extreme Learning Machine," IEEE Transactions on Wireless Communications, vol. 15, issue 2, pp. 1252–1266, 2016

[85] H. Huang, J. Yang, H. Huang, Y. Song, G. Gui, "Deep Learning for Super-Resolution Channel Estimation and DOA Estimation Based Massive MIMO System," IEEE Transactions on Vehicular Technology, vol. 67, issue 9, pp. 8549–8560, 2018

[86] G. Gui, H. Huang, Y. Song, H. Sari, "Deep Learning for an Effective Nonorthogonal Multiple Access Scheme," IEEE Transactions on Vehicular Technology, vol. 67, issue 9, pp. 8440–8450, 2018

[87] J. J. Pan, S. J. Pan, J. Yin, L. M. Ni, Q. Yang, "Tracking Mobile Users in Wireless Networks via Semi-Supervised Colocalization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, issue 3, pp. 587–600, 2012

[88] A. von Luhmann, H. Wabnitz, T. Sander, K. Muller, "M3BA: A Mobile, Modular, Multimodal Biosignal Acquisition Architecture for Miniaturized EEG-NIRS-Based Hybrid BCI and Monitoring," IEEE Transactions on Biomedical Engineering, vol. 64, issue 6, pp. 1199–1210, 2017

[89] J. Zhu, Y. Song, D. Jiang, H. Song, "A New Deep-Q-Learning-Based Transmission Scheduling Mechanism for the Cognitive Internet

of Things," IEEE Internet of Things Journal, vol. 5, issue 4, pp. 2375–2385, 2018

[90] H. Wymeersch, S. Marano, W. M. Gifford, M. Z. Win, "A Machine Learning Approach to Ranging Error Mitigation for UWB Localization," IEEE Transactions on Communications, vol. 60, issue 6, pp. 1719–1728, 2012

[91] X. Wang, L. Gao, S. Mao, "BiLoc: Bi-Modal Deep Learning for Indoor Localization With Commodity 5GHz WiFi," IEEE Access, vol. 5, issue , pp. 4209–4220, 2017

[92] M. Mozaffari, W. Saad, M. Bennis, Y. -H. Nam, M. Debbah, "A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems," IEEE Communications Surveys & Tutorials, vol. 21, issue 3, pp. 2334–2360, 2019

[93] T. Van Nguyen, Y. Jeong, H. Shin, M. Z. Win, "Machine Learning for Wideband Localization," IEEE Journal on Selected Areas in Communications, vol. 33, issue 7, pp. 1357–1380, 2015

[94] A. Krause, A. Smailagic, D. P. Siewiorek, "Context-aware mobile computing: learning context- dependent personal preferences from a wearable sensor array," IEEE Transactions on Mobile Computing, vol. 5, issue 2, pp. 113–127, 2006

[95] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, S. Pollin, "Deep Learning Models for Wireless Signal Classification With Distributed Low-Cost Spectrum Sensors," IEEE Transactions on Cognitive Communications and Networking, vol. 4, issue 3, pp. 433–445, 2018

[96] Y. Wang, K. Wu, L. M. Ni, "WiFall: Device-Free Fall Detection by Wireless Networks," IEEE Transactions on Mobile Computing, vol. 16, issue 2, pp. 581–594, 2017

[97] J. B. Predd, S. R. Kulkarni, H. V. Poor, "A Collaborative Training Algorithm for Distributed Learning," IEEE Transactions on Information Theory, vol. 55, issue 4, pp. 1856–1871, 2009

[98] X. Gao, S. Jin, C. Wen, G. Y. Li, "ComNet: Combination of Deep Learning and Expert Knowledge in OFDM Receivers," IEEE Communications Letters, vol. 22, issue 12, pp. 2627–2630, 2018

[99] P. Blasco, D. Gunduz, M. Dohler, "A Learning Theoretic Approach to Energy Harvesting Communication System Optimization," IEEE Transactions on Wireless Communications, vol. 12, issue 4, pp. 1872–1882, 2013

[100] M. Zorzi, A. Zanella, A. Testolin, M. De Filippo De Grazia, M. Zorzi, "Cognition-Based Networks: A New Perspective on Network Optimization Using Learning and Distributed Intelligence," IEEE Access, vol. 3, issue , pp. 1512–1530, 2015

[101] Yiqiang Chen, Qiang Yang, Jie Yin, Xiaoyong Chai, "Power-efficient access-point selection for indoor location estimation," IEEE Transactions on Knowledge and Data Engineering, vol. 18, issue 7, pp. 877–888, 2006

[102] C. Lin, C. Chuang, C. Huang, S. Tsai, S. Lu, Y. Chen, L. Ko, "Wireless and Wearable EEG System for Evaluating Driver Vigilance," IEEE Transactions on Biomedical Circuits and Systems, vol. 8, issue 2, pp. 165–176, 2014

[103] C. Wen, W. Shih, S. Jin, "Deep Learning for Massive MIMO CSI Feedback," IEEE Wireless Communications Letters, vol. 7, issue 5, pp. 748–751, 2018

[104] K. Bashir Shaban, A. Kadri, E. Rezk, "Urban Air Pollution Monitoring System With Forecasting Models," IEEE Sensors Journal, vol. 16, issue 8, pp. 2598–2606, 2016

[105] T. Hu, Y. Fei, "QELAR: A Machine-Learning-Based Adaptive Routing Protocol for Energy-Efficient and Lifetime-Extended Underwater Sensor Networks," IEEE Transactions on Mobile Computing, vol. 9, issue 6, pp. 796–809, 2010

[106] X. Wang, S. Wang, D. Bi, "Distributed Visual-Target-Surveillance System in Wireless Sensor Networks," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 39, issue 5, pp. 1134–1146, 2009

[107] S. M. S. Tanzil, W. Hoiles, V. Krishnamurthy, "Adaptive Scheme for Caching YouTube Content in a Cellular Network: Machine Learning Approach," IEEE Access, vol. 5, issue , pp. 5870–5881, 2017

[108] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, N. D. Sidiropoulos, "Learning to Optimize: Training Deep Neural Networks for Interference Management," IEEE Transactions on Signal Processing, vol. 66, issue 20, pp. 5438–5453, 2018

[109] X. Wang, L. Gao, S. Mao, "CSI Phase Fingerprinting for Indoor Localization With a Deep Learning Approach," IEEE Internet of Things Journal, vol. 3, issue 6, pp. 1113–1123, 2016

[110] T. J. O'Shea, T. Roy, T. C. Clancy, "Over-the-Air Deep Learning Based Radio Signal Classification," IEEE Journal of Selected Topics in Signal Processing, vol. 12, issue 1, pp. 168–179, 2018

[111] M. Chen, W. Saad, C. Yin, M. Debbah, "Echo State Networks for Proactive Caching in Cloud-Based Radio Access Networks With Mobile Users," IEEE Transactions on Wireless Communications, vol. 16, issue 6, pp. 3520–3535, 2017

[112] L. Gavrilovska, V. Atanasovski, I. Macaluso, L. A. DaSilva, "Learning and Reasoning in Cognitive Radio Networks," IEEE Communications Surveys & Tutorials, vol. 15, issue 4, pp. 1761–1777, 2013

[113] K. Merchant, S. Revay, G. Stantchev, B. Nousain, “Deep Learning for RF Device Fingerprinting in Cognitive Communication Networks,” IEEE Journal of Selected Topics in Signal Processing, vol. 12, issue 1, pp. 160–167, 2018

[114] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” IEEE Journal on Selected Areas in Communications, vol. 23, issue 2, pp. 201–220, 2005

[115] M. Kim, N. Kim, W. Lee, D. Cho, “Deep Learning-Aided SCMA,” IEEE Communications Letters, vol. 22, issue 4, pp. 720–723, 2018

[116] S. J. Pan, I. W. Tsang, J. T. Kwok, Q. Yang, “Domain Adaptation via Transfer Component Analysis,” IEEE Transactions on Neural Networks, vol. 22, issue 2, pp. 199–210, 2011

[117] B. Mager, P. Lundrigan, N. Patwari, “Fingerprint-Based Device-Free Localization Performance in Changing Environments,” IEEE Journal on Selected Areas in Communications, vol. 33, issue 11, pp. 2429–2438, 2015

[118] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, H. Kaushansky, “Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry,” IEEE Transactions on Neural Networks, vol. 11, issue 3, pp. 690–696, 2000

[119] H. Xu, W. Yu, D. Griffith, N. Golmie, “A Survey on Industrial Internet of Things: A Cyber-Physical Systems Perspective,” IEEE Access, vol. 6, issue , pp. 78238–78259, 2018

[120] Z. Liu, Z. Huang, Y. Zhou, “An Efficient Maximum Likelihood Method for Direction-of-Arrival Estimation via Sparse Bayesian Learning,” IEEE Transactions on Wireless Communications, vol. 11, issue 10, pp. 1–11, 2012

[121] M. Chen, W. Saad, C. Yin, “Echo State Networks for Self-Organizing Resource Allocation in LTE-U With Uplink-Downlink Decoupling,” IEEE Transactions on Wireless Communications, vol. 16, issue 1, pp. 3–16, 2017

[122] N. Farsad, A. Goldsmith, “Neural Network Detection of Data Sequences in Communication Systems,” IEEE Transactions on Signal Processing, vol. 66, issue 21, pp. 5663–5678, 2018

[123] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, C. S. Hong, “Caching in the Sky: Proactive Deployment of Cache-Enabled Unmanned Aerial Vehicles for Optimized Quality-of-Experience,” IEEE Journal on Selected Areas in Communications, vol. 35, issue 5, pp. 1046–1061, 2017

[124] R. He, Q. Li, B. Ai, Y. L. Geng, A. F. Molisch, V. Kristem, Z. Zhong, J. Yu, “A Kernel-Power-Density-Based Algorithm for Channel Multipath Components Clustering,” IEEE Transactions on Wireless Communications, vol. 16, issue 11, pp. 7138–7151, 2017

[125] J. Kwak, Y. Kim, L. B. Le, S. Chong, “Hybrid Content Caching in 5G Wireless Networks: Cloud Versus Edge Caching,” IEEE Transactions on Wireless Communications, vol. 17, issue 5, pp. 3030–3045, 2018

[126] M. Borgerding, P. Schniter, S. Rangan, “AMP-Inspired Deep Networks for Sparse Linear Inverse Problems,” IEEE Transactions on Signal Processing, vol. 65, issue 16, pp. 4293–4308, 2017

[127] M. Jiang, L. Hanzo, “Multiuser MIMO-OFDM for Next-Generation Wireless Systems,” Proceedings of the IEEE, vol. 95, issue 7, pp. 1430–1469, 2007

[128] G. Agamennoni, J. I. Nieto, E. M. Nebot, “Robust Inference of Principal Road Paths for Intelligent Transportation Systems,” IEEE Trans-

actions on Intelligent Transportation Systems, vol. 12, issue 1, pp. 298–308, 2011

[129] M. Vaezi, G. A. Aruma Baduge, Y. Liu, A. Arafa, F. Fang, Z. Ding, “Interplay Between NOMA and Other Emerging Technologies: A Survey,” IEEE Transactions on Cognitive Communications and Networking, vol. 5, issue 4, pp. 900–919, 2019

[130] A. Mosenia, S. Sur-Kolay, A. Raghunathan, N. K. Jha, “Wearable Medical Sensor-Based System Design: A Survey,” IEEE Transactions on Multi-Scale Computing Systems, vol. 3, issue 2, pp. 124–138, 2017

[131] F. Tang, Z. M. Fadlullah, B. Mao, N. Kato, “An Intelligent Traffic Load Prediction-Based Adaptive Channel Assignment Algorithm in SDN-IoT: A Deep Learning Approach,” IEEE Internet of Things Journal, vol. 5, issue 6, pp. 5141–5154, 2018

[132] O. B. Sezer, E. Dogdu, A. M. Ozbayoglu, “Context-Aware Computing, Learning, and Big Data in Internet of Things: A Survey,” IEEE Internet of Things Journal, vol. 5, issue 1, pp. 1–27, 2018

[133] C. Kolias, G. Kambourakis, A. Stavrou, S. Gritzalis, “Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset,” IEEE Communications Surveys & Tutorials, vol. 18, issue 1, pp. 184–208, 2016

[134] A. C. Polak, S. Dolatshahi, D. L. Goeckel, “Identifying Wireless Users via Transmitter Imperfections,” IEEE Journal on Selected Areas in Communications, vol. 29, issue 7, pp. 1469–1479, 2011

[135] G. Acampora, D. J. Cook, P. Rashidi, A. V. Vasilakos, “A Survey on Ambient Intelligence in Healthcare,” Proceedings of the IEEE, vol. 101, issue 12, pp. 2470–2494, 2013

[136] U. Satija, B. Ramkumar, M. Sabarimalai Manikandan, “Real-Time Signal Quality-Aware ECG Telemetry System for IoT-Based Health

Care Monitoring," IEEE Internet of Things Journal, vol. 4, issue 3, pp. 815–823, 2017

[137] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag, A. P. Chandrakasan, "A Micro-Power EEG Acquisition SoC With Integrated Feature Extraction Processor for a Chronic Seizure Detection System," IEEE Journal of Solid-State Circuits, vol. 45, issue 4, pp. 804–816, 2010

[138] R. R. Nadakuditi, J. W. Silverstein, "Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples," IEEE Journal of Selected Topics in Signal Processing, vol. 4, issue 3, pp. 468–480, 2010

[139] S. Marano, W. M. Gifford, H. Wymeersch, M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," IEEE Journal on Selected Areas in Communications, vol. 28, issue 7, pp. 1026–1035, 2010

[140] R. R. Yager, D. P. Filev, "Induced ordered weighted averaging operators," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 29, issue 2, pp. 141–150, 1999

[141] G. Fortino, R. Giannantonio, R. Gravina, P. Kuryloski, R. Jafari, "Enabling Effective Programming and Flexible Management of Efficient Body Sensor Network Applications," IEEE Transactions on Human-Machine Systems, vol. 43, issue 1, pp. 115–133, 2013

[142] D. Kim, G. Lee, Y. Sung, "Two-Stage Beamformer Design for Massive MIMO Downlink By Trace Quotient Formulation," IEEE Transactions on Communications, vol. 63, issue 6, pp. 2200–2211, 2015

[143] M. Haus, M. Waqas, A. Y. Ding, Y. Li, S. Tarkoma, J. Ott, "Security and Privacy in Device-to-Device (D2D) Communication: A Review,"

IEEE Communications Surveys & Tutorials, vol. 19, issue 2, pp. 1054–1079, 2017

[144] Z. Liu, Z. Huang, Y. Zhou, "Sparsity-Inducing Direction Finding for Narrowband and Wideband Signals Based on Array Covariance Vectors," IEEE Transactions on Wireless Communications, vol. 12, issue 8, pp. 1–12, 2013

[145] M. Conti, E. Sandeep Kumar, C. Lal, S. Ruj, "A Survey on Security and Privacy Issues of Bitcoin," IEEE Communications Surveys & Tutorials, vol. 20, issue 4, pp. 3416–3452, 2018

[146] M. Conti, N. Dragoni, V. Lesyk, "A Survey of Man In The Middle Attacks," IEEE Communications Surveys & Tutorials, vol. 18, issue 3, pp. 2027–2051, 2016

[147] K. Huang, N. D. Sidiropoulos, "Consensus-ADMM for General Quadratically Constrained Quadratic Programming," IEEE Transactions on Signal Processing, vol. 64, issue 20, pp. 5297–5310, 2016

[148] Y. Huang, J. Tan, Y. Liang, "Wireless Big Data: Transforming Heterogeneous Networks to Smart Networks," Journal of Communications and Information Networks, vol. 2, issue 1, pp. 19–32, 2017

[149] R. Li et al., "Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," IEEE Wireless Communications, vol. 24, no. 5, pp. 175-183, October 2017.

[150] K. Li, W. Ni, E. Tovar, A. Jamalipour, "On-Board Deep Q-Network for UAV-Assisted Online Power Transfer and Data Collection," IEEE Transactions on Vehicular Technology, vol. 68, issue 12, pp. 12215–12226, 2019

[151] Q. Wang, W. Zhang, Y. Liu, Y. Liu, "Multi-UAV Dynamic Wireless Networking With Deep Reinforcement Learning,"

[152] J. Tang, J. Song, J. Ou, J. Luo, X. Zhang, K. Wong, “Minimum Throughput Maximization for Multi-UAV Enabled WPCN: A Deep Reinforcement Learning Method,” IEEE Access, vol. 8, pp. 9124–9132, 2020

[153] T. Zhang, Z. Wang, Y. Liu, W. Xu, A. Nallanathan, “Caching Placement and Resource Allocation for Cache-Enabling UAV NOMA Networks,” IEEE Transactions on Vehicular Technology, vol. 69, issue 11, pp. 12897–12911, 2020

[154] J. Chen, F. Yan, S. Mao, F. Shen, W. Xia, Y. Wu, L. Shen, “Efficient Data Collection in Large-Scale UAV-aided Wireless Sensor Networks,” 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–5, 2019

[155] X. Du, H. Van Nguyen, C. Jiang, Y. Li, F. R. Yu, Z. Han, “Virtual Relay Selection in LTE-V: A Deep Reinforcement Learning Approach to Heterogeneous Data,” IEEE Access, vol. 8, pp. 102477–102492, 2020

[156] L. U. Khan, I. Yaqoob, M. Imran, Z. Han, C. S. Hong, “6G Wireless Systems: A Vision, Architectural Elements, and Future Directions,” IEEE Access, vol. 8, pp. 147029–147044, 2020

[157] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, D. I. Kim, “Applications of Deep Reinforcement Learning in Communications and Networking: A Survey,” IEEE Communications Surveys & Tutorials, vol. 21, issue 4, pp. 3133–3174, 2019

[158] J. Li, H. Gao, T. Lv, Y. Lu, “Deep reinforcement learning based computation offloading and resource allocation for MEC,” 2018 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, 2018

[159] J. Zhu, Y. Song, D. Jiang, H. Song, “A New Deep-Q-Learning-Based Transmission Scheduling Mechanism for the Cognitive Internet

of Things," IEEE Internet of Things Journal, vol. 5, issue 4, pp. 2375–2385, 2018

[160] C. Qiu, F. R. Yu, H. Yao, C. Jiang, F. Xu, C. Zhao, "Blockchain-Based Software-Defined Industrial Internet of Things: A Dueling Deep Q -Learning Approach," IEEE Internet of Things Journal, vol. 6, issue 3, pp. 4627–4639, 2019

[161] H. He, H. Shan, A. Huang, Q. Ye, W. Zhuang, "Reinforcement Learning-Based Computing and Transmission Scheduling for LTE-U-Enabled IoT," 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2018

[162] J. Hribar, A. Marinescu, G. A. Ropokis, L. A. DaSilva, "Using Deep Q-Learning to Prolong the Lifetime of Correlated Internet of Things Devices," 2019 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1–6, 2019

[163] K. Li, W. Ni, E. Tovar, A. Jamalipour, "Deep Q-Learning based Resource Management in UAV-assisted Wireless Powered IoT Networks," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), pp. 1–6, 2020

[164] Z. Shi, X. Xie, H. Lu, H. Yang, M. Kadoch, M. Cheriet, "Deep-Reinforcement-Learning-Based Spectrum Resource Management for Industrial Internet of Things," IEEE Internet of Things Journal, vol. 8, issue 5, pp. 3476–3489, 2021

[165] F. Jiang, Z. Yuan, C. Sun, J. Wang, "Deep Q-Learning-Based Content Caching With Update Strategy for Fog Radio Access Networks," IEEE Access, vol. 7, pp. 97505–97514, 2019

[166] Y. Fang, J. Xiong, P. Cheng, W. Zhang, "Distributed Caching Popular Services by Using Deep Q-Learning in Converged Networks," 2019

IEEE 90th Vehicular Technology Conference (VTC2019-Fall), pp. 1–5, 2019

[167] M. Chen, W. Saad, C. Yin, "Deep Learning for 360° Content Transmission in UAV-Enabled Virtual Reality," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1–6, 2019

[168] K. Zhang, S. Leng, X. Peng, L. Pan, S. Maharjan, Y. Zhang, "Artificial Intelligence Inspired Transmission Scheduling in Cognitive Vehicular Communications and Networks," IEEE Internet of Things Journal, vol. 6, issue 2, pp. 1987–1997, 2019

[169] Y. S. Nasir, D. Guo, "Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks," IEEE Journal on Selected Areas in Communications, vol. 37, issue 10, pp. 2239–2250, 2019

[170] M. Gadaleta, F. Chiariotti, M. Rossi, A. Zanella, "D-DASH: A Deep Q-Learning Framework for DASH Video Streaming," IEEE Transactions on Cognitive Communications and Networking, vol. 3, issue 4, pp. 703–718, 2017

[171] T. Yang, Y. Hu, M. C. Gursoy, A. Schmeink, R. Mathar, "Deep Reinforcement Learning based Resource Allocation in Low Latency Edge Computing Networks," 2018 15th International Symposium on Wireless Communication Systems (ISWCS), pp. 1–5, 2018

[172] C. Qiu, H. Yao, F. R. Yu, F. Xu, C. Zhao, "Deep Q-Learning Aided Networking, Caching, and Computing Resources Allocation in Software-Defined Satellite-Terrestrial Networks," IEEE Transactions on Vehicular Technology, vol. 68, issue 6, pp. 5871–5883, 2019

[173] F. Meng, P. Chen, L. Wu, "Power Allocation in Multi-User Cellular Networks with Deep Q Learning Approach," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1–6, 2019

[174] M. Chen, W. Saad, C. Yin, "Echo-Liquid State Deep Learning for 360° Content Transmission and Caching in Wireless VR Networks With Cellular-Connected UAVs," IEEE Transactions on Communications, vol. 67, issue 9, pp. 6386–6400, 2019

[175] L. Hou, L. Lei, K. Zheng, X. Wang, "A Q -Learning-Based Proactive Caching Strategy for Non-Safety Related Services in Vehicular Networks," IEEE Internet of Things Journal, vol. 6, issue 3, pp. 4512–4520, 2019

[176] Y. He, C. Liang, F. R. Yu, Z. Han, "Trust-Based Social Networks with Computing, Caching and Communications: A Deep Reinforcement Learning Approach," IEEE Transactions on Network Science and Engineering, vol. 7, issue 1, pp. 66–79, 2020

[177] T. Oda, R. Obukata, M. Ikeda, L. Barolli, M. Takizawa, "Design and Implementation of a Simulation System Based on Deep Q-Network for Mobile Actor Node Control in Wireless Sensor and Actor Networks," 2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp. 195–200, 2017

[178] L. Li, Y. Xu, J. Yin, W. Liang, X. Li, W. Chen, Z. Han, "Deep Reinforcement Learning Approaches for Content Caching in Cache-Enabled D2D Networks," IEEE Internet of Things Journal, vol. 7, issue 1, pp. 544–557, 2020

[179] Y. Su, X. Lu, Y. Zhao, L. Huang, X. Du, "Cooperative Communications With Relay Selection Based on Deep Reinforcement Learning in Wireless Sensor Networks," IEEE Sensors Journal, vol. 19, issue 20, pp. 9561–9569, 2019

[180] T. T. Anh, N. C. Luong, D. Niyato, D. I. Kim, L. -C. Wang, "Efficient Training Management for Mobile Crowd-Machine Learning: A Deep Reinforcement Learning Approach," IEEE Wireless Communications Letters, vol. 8, issue 5, pp. 1345–1348, 2019

[181] N. Van Huynh, D. N. Nguyen, D. T. Hoang, E. Dutkiewicz, “Jam Me If You Can: Defeating Jammer With Deep Dueling Neural Network Architecture and Ambient Backscattering Augmented Communications,” IEEE Journal on Selected Areas in Communications, vol. 37, issue 11, pp. 2603–2620, 2019

[182] B. Peng, G. Seco-Granados, E. Steinmetz, M. Frohle, H. Wymeersch, “Decentralized Scheduling for Cooperative Localization With Deep Reinforcement Learning,” IEEE Transactions on Vehicular Technology, vol. 68, issue 5, pp. 4295–4305, 2019

[183] R. Ali, N. Shahin, Y. B. Zikria, B. Kim, S. W. Kim, “Deep Reinforcement Learning Paradigm for Performance Optimization of Channel Observation-Based MAC Protocols in Dense WLANs,” IEEE Access, vol. 7, pp. 3500–3511, 2019

[184] F. B. Mismar, B. L. Evans, A. Alkhateeb, “Deep Reinforcement Learning for 5G Networks: Joint Beamforming, Power Control, and Interference Coordination,” IEEE Transactions on Communications, vol. 68, issue 3, pp. 1581–1592, 2020

[185] A. M. Koushik, F. Hu, S. Kumar, “Deep Q -Learning-Based Node Positioning for Throughput-Optimal Communications in Dynamic UAV Swarm Network,” IEEE Transactions on Cognitive Communications and Networking, vol. 5, issue 3, pp. 554–566, 2019

[186] X. Fu, F. R. Yu, J. Wang, Q. Qi, J. Liao, “Dynamic Service Function Chain Embedding for NFV-Enabled IoT: A Deep Reinforcement Learning Approach,” IEEE Transactions on Wireless Communications, vol. 19, issue 1, pp. 507–519, 2020 IEEE Communications Letters, vol. 23, issue 12, pp. 2243–2246, 2019

[187] J. Liu, X. Tao, J. Lu, “QoE-Oriented Rate Adaptation for DASH With Enhanced Deep Q-Learning,” IEEE Access, vol. 7, pp. 8454–8469, 2019

[188] M. Kozy, J. Yu, R. M. Buehrer, A. Martone, K. Sherbondy, "Applying Deep-Q Networks to Target Tracking to Improve Cognitive Radar," 2019 IEEE Radar Conference (RadarConf), pp. 1–6, 2019

[189] M. Li, X. Zhao, H. Liang, F. Hu, "Deep Reinforcement Learning Optimal Transmission Policy for Communication Systems With Energy Harvesting and Adaptive MQAM," IEEE Transactions on Vehicular Technology, vol. 68, issue 6, pp. 5782–5793, 2019

[190] M. Elsayed, M. Erol-Kantarci, "Deep Reinforcement Learning for Reducing Latency in Mission Critical Services," 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2018

[191] Y. He, C. Liang, F. R. Yu, V. C. M. Leung, "Integrated Computing, Caching, and Communication for Trust-Based Social Networks: A Big Data DRL Approach," 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2018

[192] C. Luo, J. Ji, Q. Wang, L. Yu, P. Li, "Online Power Control for 5G Wireless Communications: A Deep Q-Network Approach," 2018 IEEE International Conference on Communications (ICC), pp. 1–6, 2018

[193] F. Xu, F. Yang, S. Bao, C. Zhao, "DQN Inspired Joint Computing and Caching Resource Allocation Approach for Software Defined Information-Centric Internet of Things Network," IEEE Access, vol. 7, pp. 61987–61996, 2019

[194] L. Zhang, Z. Zhang, C. Huang, H. Deng, H. Lin, B. Tseng, J. Drewniak, C. Hwang, "Decoupling Capacitor Selection Algorithm for PDN Based on Deep Reinforcement Learning," 2019 IEEE International Symposium on Electromagnetic Compatibility, Signal & Power Integrity (EMC+SIPI), pp. 616–620, 2019

[195] H. Huang, Y. Yang, H. Wang, Z. Ding, H. Sari, F. Adachi, "Deep Reinforcement Learning for UAV Navigation Through Massive MIMO

Technique," IEEE Transactions on Vehicular Technology, vol. 69, issue 1, pp. 1117–1121, 2020

[196] Z. Zhang, H. Chen, M. Hua, C. Li, Y. Huang, L. Yang, "Double Coded Caching in Ultra Dense Networks: Caching and Multicast Scheduling via Deep Reinforcement Learning," IEEE Transactions on Communications, vol. 68, issue 2, pp. 1071–1086, 2020

[197] F. B. Mismar, B. L. Evans, "Deep Q-Learning for Self-Organizing Networks Fault Management and Radio Performance Improvement," 2018 52nd Asilomar Conference on Signals, Systems, and Computers, pp. 1457–1461, 2018

[198] M. Elsayed, M. Erol-Kantarci, "Deep Q-Learning for Low-Latency Tactile Applications: Microgrid Communications," 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), pp. 1–6, 2018

[199] Y. Li, W. Zhang, C. Wang, J. Sun, Y. Liu, "Deep Reinforcement Learning for Dynamic Spectrum Sensing and Aggregation in Multi-Channel Wireless Networks," IEEE Transactions on Cognitive Communications and Networking, vol. 6, issue 2, pp. 464–475, 2020

[200] C. Qiu, F. R. Yu, F. Xu, H. Yao, C. Zhao, "Permissioned Blockchain-Based Distributed Software-Defined Industrial Internet of Things," 2018 IEEE Globecom Workshops (GC Wkshps), pp. 1–7, 2018

[201] T. Li, X. Zhu, X. Liu, "An End-to-End Network Slicing Algorithm Based on Deep Q-Learning for 5G Network," IEEE Access, vol. 8, pp. 122229–122240, 2020

[202] A. Zhu, S. Guo, M. Ma, H. Feng, B. Liu, X. Su, M. Guo, Q. Jiang, "Computation Offloading for Workflow in Mobile Edge Computing Based on Deep Q-Learning," 2019 28th Wireless and Optical Communications Conference (WOCC), pp. 1–5, 2019

[203] C. Zhang, M. Dong, K. Ota, "Fine-Grained Management in 5G: DQL Based Intelligent Resource Allocation for Network Function Virtualization in C-RAN," IEEE Transactions on Cognitive Communications and Networking, vol. 6, issue 2, pp. 428–435, 2020

[204] A. El-Amine, M. Iturralde, H. A. Haj Hassan, L. Nuaymi, "A Distributed Q-Learning Approach for Adaptive Sleep Modes in 5G Networks," 2019 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, 2019

[205] Y. Zhao, J. Hu, K. Yang, S. Cui, "Deep Reinforcement Learning Aided Intelligent Access Control in Energy Harvesting Based WLAN," IEEE Transactions on Vehicular Technology, vol. 69, issue 11, pp. 14078–14082, 2020

[206] Z. Zou, R. Yin, X. Chen, C. Wu, "Deep Reinforcement Learning for D2D transmission in unlicensed bands," 2019 IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops), pp. 42–47, 2019

[207] M. E. Morocho-Cayamcela, H. Lee, W. Lim, "Machine Learning to Improve Multi-Hop Searching and Extended Wireless Reachability in V2X," IEEE Communications Letters, vol. 24, issue 7, pp. 1477–1481, 2020

[208] Z. Zhu, Z. Zhang, W. Yan, Y. Huang, L. Yang, "Proactive Caching in Auto Driving Scene via Deep Reinforcement Learning," 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6, 2019

[209] C. Sun, Z. Shi, F. Jiang, "A Machine Learning Approach for Beamforming in Ultra Dense Network Considering Selfish and Altruistic Strategy," IEEE Access, vol. 8, pp. 6304–6315, 2020

[210] Z. Zhang, Y. Zheng, C. Li, Y. Huang, L. Yang, "Cache-Enabled Adaptive Bit Rate Streaming via Deep Self-Transfer Reinforcement Learning," 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6, 2018

[211] T. Zhang, C. Xu, B. Zhang, X. Kuang, Y. Wang, S. Yang, G. Muntean, "DQ-RM: Deep Reinforcement Learning-based Route Mutation Scheme for Multimedia Services," 2020 International Wireless Communications and Mobile Computing (IWCMC), pp. 291–296, 2020

[212] F. Tang, Y. Zhou, N. Kato, "Deep Reinforcement Learning for Dynamic Uplink/Downlink Resource Allocation in High Mobility 5G HetNet," IEEE Journal on Selected Areas in Communications, vol. 38, issue 12, pp. 2773–2782, 2020

[213] A. Iqbal, M. -L. Tham, Y. C. Chang, "Double Deep Q-Network for Power Allocation in Cloud Radio Access Network," 2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology (CCET), pp. 272–277, 2020

[214] A. Iqbal, M. -L. Tham, Y. C. Chang, "Double Deep Q-Network-Based Energy-Efficient Resource Allocation in Cloud Radio Access Network," IEEE Access, vol. 9, pp. 20440–20449, 2021

[215] F. Li, Y. Zhu, Y. Xu, "Dynamic Multi-channel Access in Wireless System with Deep Reinforcement Learning," 2020 12th International Conference on Advanced Computational Intelligence (ICACI), pp. 283–287, 2020

[216] H. He, H. Shan, A. Huang, Q. Ye, W. Zhuang, "Edge-Aided Computing and Transmission Scheduling for LTE-U-Enabled IoT," IEEE Transactions on Wireless Communications, vol. 19, issue 12, pp. 7881–7896, 2020

[217] C. E. Thornton, R. M. Buehrer, A. F. Martone, K. D. Sherbondy, "Experimental Analysis of Reinforcement Learning Techniques for Spectrum Sharing Radar," 2020 IEEE International Radar Conference (RADAR), pp. 67–72, 2020

[218] X. Han, J. Wang, J. Xue, Q. Zhang, "Intelligent Decision-Making for 3-Dimensional Dynamic Obstacle Avoidance of UAV Based on Deep Reinforcement Learning," 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6, 2019

[219] F. Jameel, M. A. Jamshed, Z. Chang, R. Jantti, H. Pervaiz, "Low Latency Ambient Backscatter Communications with Deep Q-Learning for Beyond 5G Applications," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), pp. 1–6, 2020

[220] A. Chandramohan, M. Poel, B. Meijerink, G. Heijenk, "Machine Learning for Cooperative Driving in a Multi-Lane Highway Environment," 2019 Wireless Days (WD), pp. 1–4, 2019

[221] S. R. Pokhrel, S. Garg, "Multipath Communication With Deep Q-Network for Industry 4.0 Automation and Orchestration," IEEE Transactions on Industrial Informatics, vol. 17, issue 4, pp. 2852–2859, 2021

[222] F. Meng, P. Chen, L. Wu, J. Cheng, "Power Allocation in Multi-User Cellular Networks: Deep Reinforcement Learning Approaches," IEEE Transactions on Wireless Communications, vol. 19, issue 10, pp. 6255–6267, 2020

[223] K. Wang, Y. Zhou, Y. Yang, X. Yuan, X. Luo, "Task Offloading in NOMA-Based Fog Computing Networks: A Deep Q-Learning Approach," 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2019

[224] J. Guo, Y. Huo, X. Shi, J. Wu, P. Yu, L. Feng, W. Li, "3D Aerial Vehicle Base Station (UAV-BS) Position Planning based on Deep Q-Learning

for Capacity Enhancement of Users With Different QoS Requirements," 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 1508–1512, 2019

[225] Q. Huang, M. Kadoch, "5G Resource Scheduling for Low-latency Communication: A Reinforcement Learning Approach," 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), pp. 1–5, 2020

[226] F. D. Kronewitter, S. Lee, K. Oliphant, "A Cognitive ML Agent for Airborne Networking," MILCOM 2019 - 2019 IEEE Military Communications Conference (MILCOM), pp. 115–120, 2019

[227] M. Wu, W. Huang, K. Sun, H. Zhang, "A DQN-Based Handover Management for SDN-Enabled Ultra-Dense Networks," 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), pp. 1–6, 2020

[228] P. K. H. Nguyen, V. H. Nguyen, V. L. Do, "A Deep Double-Q Learning-based Scheme for Anti-Jamming Communications," 2020 28th European Signal Processing Conference (EUSIPCO), pp. 1566–1570, 2021

[229] X. Gao, Z. Dou, L. Qi, "A New Distributed Dynamic Spectrum Access Model Based on DQN," 2020 15th IEEE International Conference on Signal Processing (ICSP), vol. 1, pp. 351–355, 2020

[230] S. Tomovic, I. Radusinovic, "A Novel Deep Q-learning Method for Dynamic Spectrum Access," 2020 28th Telecommunications Forum (TELFOR), pp. 1–4, 2020

[231] E. Balevi, J. G. Andrews, "A Novel Deep Reinforcement Learning Algorithm for Online Antenna Tuning," 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2019

[232] Z. Shi, X. Xie, M. Kadoch, M. Cheriet, "A Spectrum Resource Sharing Algorithm for IoT Networks based on Reinforcement Learning," 2020 International Wireless Communications and Mobile Computing (IWCMC), pp. 1019–1024, 2020

[233] X. Han, W. Yang, M. Li, L. Zhang, F. Yan, X. Ma, "A Trajectory Planning Algorithm for Data Collection in UAV-aided Wireless Sensor Networks," 2020 IEEE 6th International Conference on Computer and Communications (ICCC), pp. 846–851, 2020

[234] Q. V. Do, I. Koo, "A Transfer Deep Q-Learning Framework for Resource Competition in Virtual Mobile Networks With Energy-Harvesting Base Stations," IEEE Systems Journal, vol. 15, issue 1, pp. 319–330, 2021

[235] O. Bouhamed, H. Ghazzai, H. Besbes, Y. Massoud, "A UAV-Assisted Data Collection for Wireless Sensor Networks: Autonomous Navigation and Scheduling," IEEE Access, vol. 8, pp. 110446–110460, 2020

[236] A. Karapantelakis, E. Fersman, "A deep-Q learning approach to mobile operator collaboration," Journal of Communications and Networks, vol. 22, issue 6, pp. 455–466, 2020

[237] V. Sadhu, C. Sun, A. Karimian, R. Tron, D. Pompili, "Aerial-DeepSearch: Distributed Multi-Agent Deep Reinforcement Learning for Search Missions," 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), pp. 165–173, 2020

[238] Z. Hu, S. Cong, T. Song, K. Bian, L. Song, "AirScope: Mobile Robots-Assisted Cooperative Indoor Air Quality Sensing by Distributed Deep Reinforcement Learning," IEEE Internet of Things Journal, vol. 7, issue 9, pp. 9189–9200, 2020

[239] L. Zhang, W. Huang, J. Juang, H. Lin, B. -C. Tseng, C. Hwang, "An Enhanced Deep Reinforcement Learning Algorithm for Decoupling Capacitor Selection in Power Distribution Network Design," 2020 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI), pp. 245–250, 2020

[240] Y. Al-Eryani, M. Akrout, E. Hossain, "Antenna Clustering for Simultaneous Wireless Information and Power Transfer in a MIMO Full-Duplex System: A Deep Reinforcement Learning-Based Design," IEEE Transactions on Communications, vol. 69, issue 4, pp. 2331–2345, 2021

[241] S. Zeng, Y. Ren, Y. Wang, T. Zhao, Z. Qian, "Caching Strategy Based on Deep Q-Learning in Device-to-device Scenario," 2019 12th International Symposium on Computational Intelligence and Design (ISCID), vol. 1, pp. 175–179, 2019

[242] S. Wu, Y. Wang, L. Bai, "Deep Convolutional Neural Network Assisted Reinforcement Learning Based Mobile Network Power Saving," IEEE Access, vol. 8, pp. 93671–93681, 2020

[243] C. Sun, J. Zhou, X. Zhou, X. Zhang, W. Wang, "Deep Learning Enabled Dynamic Reactive Video Caching in Mobile Edge Networks," 2018 IEEE International Conference on Communication Systems (ICCS), pp. 280–285, 2018

[244] L. Wang, K. Wang, C. Pan, X. Chen, N. Aslam, "Deep Q-Network Based Dynamic Trajectory Design for UAV-Aided Emergency Communications," Journal of Communications and Information Networks, vol. 5, issue 4, pp. 393–402, 2020

[245] T. Lin, Z. Su, Q. Xu, R. Xing, D. Fang, "Deep Q-Network Based Energy Scheduling in Retail Energy Market," IEEE Access, vol. 8, pp. 69284–69295, 2020

[246] D. Shome, A. Kudeshia, "Deep Q-learning for 5G network slicing with diverse resource stipulations and dynamic data traffic," 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 134–139, 2021

[247] C. Huang, G. Chen, Y. Gong, P. Xu, "Deep Reinforcement Learning Based Relay Selection in Delay-Constrained Secure Buffer-Aided

CRNs," GLOBECOM 2020 - 2020 IEEE Global Communications Conference, pp. 1–6, 2020

[248] C. E. Thornton, M. A. Kozy, R. M. Buehrer, A. F. Martone, K. D. Sherbondy, "Deep Reinforcement Learning Control for Radar Detection and Tracking in Congested Spectral Environments," IEEE Transactions on Cognitive Communications and Networking, vol. 6, issue 4, pp. 1335–1349, 2020

[249] H. H. Esmat, B. Lorenzo, "Deep Reinforcement Learning based Dynamic Edge/Fog Network Slicing," GLOBECOM 2020 - 2020 IEEE Global Communications Conference, pp. 1–6, 2020

[250] Z. Hu, T. Song, K. Bian, L. Song, "Deep Reinforcement Learning based Indoor Air Quality Sensing by Cooperative Mobile Robots," 2020 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, 2020

[251] Z. Hajiakhondi-Meybodi, A. Mohammadi, J. Abouei, "Deep Reinforcement Learning for Trustworthy and Time-Varying Connection Scheduling in a Coupled UAV-Based Femtocaching Architecture," IEEE Access, vol. 9, pp. 32263–32281, 2021

[252] Y. Peng, L. Liu, Y. Zhou, J. Shi, J. Li, "Deep Reinforcement Learning-Based Dynamic Service Migration in Vehicular Networks," 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2019

[253] Y. Wang, Y. Li, T. Lan, V. Aggarwal, "DeepChunk: Deep Q-Learning for Chunk-Based Caching in Wireless Data Processing Networks," IEEE Transactions on Cognitive Communications and Networking, vol. 5, issue 4, pp. 1034–1045, 2019

[254] N. V. Huynh, D. N. Nguyen, D. T. Hoang, E. Dutkiewicz, M. Mueck, S. Srikanteswara, "Defeating Jamming Attacks with Ambient Backscat-

ter Communications," 2020 International Conference on Computing, Networking and Communications (ICNC), pp. 405–409, 2020

[255] T. H. Dinh, M. A. Alsheikh, S. Gong, D. Niyato, Z. Han, Y. Liang, "Defend Jamming Attacks: How to Make Enemies Become Friends," 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2019

[256] U. Kaytaz, S. Ucar, B. Akgun, S. Coleri, "Distributed Deep Reinforcement Learning with Wideband Sensing for Dynamic Spectrum Access," 2020 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, 2020

[257] S. Jiang, Y. Chang, K. Fukawa, "Distributed Inter-cell Interference Coordination for Small Cell Wireless Communications: A Multi-Agent Deep Q-Learning Approach," 2020 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 1–5, 2020

[258] S. Bhardwaj, J. -M. Lee, D. -S. Kim, "Double Deep Q-Learning Based Channel Estimation for Industrial Wireless Networks," 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1318–1320, 2020

[259] Y. Lin, H. Gao, W. Xu, Y. Lu, "Dynamic Antenna Configuration for 3D Massive MIMO System via Deep Reinforcement Learning," 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–6, 2020

[260] S. Liu, J. Wu, J. He, "Dynamic Multichannel Sensing in Cognitive Radio: Hierarchical Reinforcement Learning," IEEE Access, vol. 9, pp. 25473–25481, 2021

[261] H. T. Huong Giang, P. Duy Thanh, I. Koo, "Dynamic Power Allocation Scheme for NOMA Uplink in Cognitive Radio Networks Using

Deep Q Learning," 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 137–142, 2020

[262] S. Chen, L. Li, Z. Chen, S. Li, "Dynamic Pricing for Smart Mobile Edge Computing: A Reinforcement Learning Approach," IEEE Wireless Communications Letters, vol. 10, issue 4, pp. 700–704, 2021

[263] Z. Chang, X. Zhou, Z. Wang, H. Li, X. Zhang, "Edge-assisted Adaptive Video Streaming with Deep Learning in Mobile Edge Networks," 2019 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, 2019

[264] M. Roudneshin, A. M. M. Sizkouhi, A. G. Aghdam, "Effective Learning Algorithms for Search and Rescue Missions in Unknown Environments," 2019 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE), pp. 76–80, 2019

[265] D. Shi, F. Tian, S. Wu, "Energy Efficiency Optimization in Heterogeneous Networks Based on Deep Reinforcement Learning," 2020 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1–6, 2020

[266] A. Ashiquzzaman, H. Lee, T. Um, J. Kim, "Energy-Efficient IoT Sensor Calibration With Deep Reinforcement Learning," IEEE Access, vol. 8, pp. 97045–97055, 2020

[267] Q. Zhang, J. Miao, Z. Zhang, F. R. Yu, F. Fu, T. Wu, "Energy-Efficient Video Streaming in UAV-Enabled Wireless Networks: A Safe-DQN Approach," GLOBECOM 2020 - 2020 IEEE Global Communications Conference, pp. 1–7, 2020

[268] J. Li, T. Yang, H. Feng, "Intelligent Maritime Communications Enabled by Deep Reinforcement Learning," 2019 IEEE/CIC International Conference on Communications in China (ICCC), pp. 786–791, 2019

[269] M. G. Khoshkholgh, H. Yanikomeroglu, "Learning Power Control From a Fixed Batch of Data," IEEE Wireless Communications Letters, vol. 10, issue 3, pp. 512–516, 2021

[270] Z. Cheng, M. LiWang, N. Chen, H. Lin, Z. Gao, L. Huang, "Learning-Based Joint User-AP Association and Resource Allocation in Ultra Dense Network," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), pp. 1–5, 2020

[271] M. Chen, Y. Xiao, Q. Li, K. Chen, "Minimizing Age-of-Information for Fog Computing-supported Vehicular Networks with Deep Q-learning," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), pp. 1–6, 2020

[272] G. Hao, W. Ni, H. Tian, L. Cao, "Mobility-Aware Trajectory Design for Aerial Base Station Using Deep Reinforcement Learning," 2020 International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1131–1136, 2020

[273] X. Zhang, W. Lin, Y. Li, Q. Cui, X. Tao, X. Huang, P. Ren, "Moving Server: Follow-up Computation Offloading Paradigm for Vehicular Users," 2020 IEEE/CIC International Conference on Communications in China (ICCC), pp. 226–231, 2020

[274] X. Wang, M. Cenk Gursoy, "Multi-Agent Double Deep Q-Learning for Beamforming in mmWave MIMO Networks," 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–6, 2020

[275] Z. Luo, Q. Chen, G. Yu, "Multi-Agent Reinforcement Learning Based Unlicensed Resource Sharing for LTE-U Networks," 2018 IEEE International Conference on Communication Systems (ICCS), pp. 427–432, 2018

[276] X. Han, J. Wang, Q. Zhang, X. Qin, M. Sun, "Multi-UAV Automatic Dynamic Obstacle Avoidance with Experience-shared A2C," 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 330–335, 2019

[277] S. Liang, H. Wan, T. Qin, J. Li, W. Chen, "Multi-user Computation Offloading for Mobile Edge Computing: A Deep Reinforcement Learning and Game Theory Approach," 2020 IEEE 20th International Conference on Communication Technology (ICCT), pp. 1534–1539, 2020

[278] Y. Su, M. Liwang, Z. Gao, L. Huang, X. Du, M. Guizani, "Optimal Cooperative Relaying and Power Control for IoUT Networks With Reinforcement Learning," IEEE Internet of Things Journal, vol. 8, issue 2, pp. 791–801, 2021

[279] P. Yan, S. Choudhury, "Optimizing Mobile Edge Computing Multi-Level Task Offloading via Deep Reinforcement Learning," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), pp. 1–7, 2020

[280] W. Guo, "Partially Explainable Big Data Driven Deep Reinforcement Learning for Green 5G UAV," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), pp. 1–7, 2020

[281] Z. Zhang, R. Wang, F. R. Yu, F. Fu, Q. Yan, Q. Jiao, "QoE Aware Transcoding for Live Streaming in SDN-Based Cloud-Aided HetNets: An Actor-Critic Approach," 2019 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1–6, 2019

[282] W. Ahsan, W. Yi, Y. Liu, Z. Qin, A. Nallanathan, "Reinforcement Learning for User Clustering in NOMA-Enabled Uplink IoT," 2020 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1–6, 2020

[283] X. He, “Reinforcement-Learning-Based Solutions to Power Issues in Wireless IoT System,” 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), pp. 174–178, 2020

[284] P. Luong, F. Gagnon, F. Labeau, “Resource Allocation in UAV-Assisted Wireless Networks Using Reinforcement Learning,” 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), pp. 1–6, 2020

[285] M. Mohsenivatani, M. Darabi, S. Parsaeefard, M. Ardebilipour, B. Maham, “Throughput Maximization in C-RAN Enabled Virtualized Wireless Networks via Multi-Agent Deep Reinforcement Learning,” 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–6, 2020

[286] Z. Xiong, Y. Zhang, W. Y. B. Lim, J. Kang, D. Niyato, C. Leung, C. Miao, “UAV-Assisted Wireless Energy and Data Transfer With Deep Reinforcement Learning,” IEEE Transactions on Cognitive Communications and Networking, vol. 7, issue 1, pp. 85–99, 2021

[287] I. M. Braga, E. d. O. Cavalcante, G. Fodor, Y. C. B. Silva, C. F. M. e Silva, W. C. Freitas, “User Scheduling Based on Multi-Agent Deep Q-Learning for Robust Beamforming in Multicell MISO Systems,” IEEE Communications Letters, vol. 24, issue 12, pp. 2809–2813, 2020

[288] C. Zhang, M. Dong, K. Ota, “Vehicular Multi-slice Optimization in 5G: Dynamic Preference Policy using Reinforcement Learning,” GLOBECOM 2020 - 2020 IEEE Global Communications Conference, pp. 1–6, 2020

[289] S. Maghsudi and E. Hossain, “Multi-armed bandits with application to 5G small cells,” IEEE Wireless Communications, vol. 23, no. 3, pp. 64-73, 2016.

[290] T. M. Cover and J. A. Thomas, “Elements of Information Theory,”. New York: Wiley, 1991.

[291] J. Song, M. Sheng, T. Q. S. Quek, C. Xu, X. Wang, “Learning-Based Content Caching and Sharing for Wireless Networks,” IEEE Transactions on Communications, vol. 65, issue 10, pp. 4309–4324, 2017

[292] Q. Wang, P. Xu, K. Ren, X. Li, “Towards Optimal Adaptive UFH-Based Anti-Jamming Wireless Communication,” IEEE Journal on Selected Areas in Communications, vol. 30, issue 1, pp. 16–30, 2012

[293] V. Va, T. Shimizu, G. Bansal, R. W. Heath, “Online Learning for Position-Aided Millimeter Wave Beam Training,” IEEE Access, vol. 7, pp. 30507–30526, 2019

[294] ns-3 : a discrete-event network simulator for Internet systems, `https://www.nsnam.org/`

[295] “Beyond 5G/6G White Paper,” National Institute of Information and Communications Technology, 2021 `https://www2.nict.go.jp/idi/`

[296] “BBR congestion control,” Google Inc., `https://github.com/google/bbr/blob/master/Presentations/bbr-2017-02-08-google-net-research-summit.pdf`

[297] S. Takeuchi, M. Hasegawa, K. Kanno, et al., “Dynamic channel selection in wireless communications via a multi-armed bandit algorithm using laser chaos time series,” Scientific Report 10, 1574, 2020.

[298] J. Ma, S. Hasegawa, S.-J. Kim, M. Hasegawa, “A Reinforcement-Learning-Based Distributed Resource Selection Algorithm for Massive IoT,” Applied Science, 2019, vo. 9, 3730.

[299] H. Kanemasa, A. Li, M. Naruse, N. Chauvet and M. Hasegawa, “Dynamic Channel Bonding Using Laser Chaos Decision Maker in

WLANs," 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2021, pp. 078–082,

[300] Z. Duan, N. Okada, A. Li, M. Naruse, N. Chauvet and M. Hasegawa, "High-speed Optimization of User Pairing in NOMA System Using Laser Chaos Based MAB Algorithm," 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2021, pp. 073–077.

[301] K. Oshima, T. Onishi, S-J. Kim, J. Ma, M. Hasegawa, "Efficient wireless network selection by using multi-armed bandit algorithm for mobile terminals," Nonlinear Theory and Its Applications, IEICE, vol. 11, Issue 1, pp. 68–77, 2020.

[302] K. Oshima, T. Kobayashi, Y. Taenaka, K. Kuroda, M. Hasegawa, "Wireless network optimization method based on cognitive cycle using machine learning," IEICE ComEX, vol. 7, no. 7, pp. 278–283, 2018.

[303] K. Oshima, T. Kobayashi, Y. Taenaka, K. Kuroda, M. Hasegawa, "Autonomous Wireless System Optimization Method based on Cross-layer Modeling using Machine Learning," IEEE International Conference on Ubiquitous and Future Networks, pp. 239–244, 2019.

Acknowledgments

To finish this dissertation, the fact that this dissertation would not have been possible without many people. First of all, I owe much gratitude to my supervisor Prof. Mikio Hasegawa. He alway gave me a fruitful discussions, lights to the darkness, and courage to complete my journey for Ph.D. I would also particularly like to thank Prof. Takayuki Hamamoto, Prof. Takahiro Yoshida, Prof. Yuzur Ueda, Prof. Hiroyuki Yashima, and Prof. Makoto Itami for meaningful discussion for my research and dissertation.

With my colleagues in Tokyo University of Science, laboratory members led by Prof. Hasegawa, I always had wonderful time during my Ph.D. I am glad to be a member of the laboratory.

To my colleagues in Kozo Keikaku Engineering Inc., they alwasy supported and encouraged me. It is for sure that their kindness helped me to achieve this goal.

Finally, to my family. I cannot express how I am grateful to my wife, Yuka. She opened my road to Ph.D., always supported me, fostering our lovely angels: Kano and Kohei. My parents and sisters and brothers are always undoubt allies of me. And, to my uncle Dr. Yoshio Murata, might see me from heaven. He said to me that to pursue higher top of the mountain helps you to broaden your insight to the world. I completely agree with him.

I would thank all of my friends wishing me to achieve the goal. They are treasure in my life.

2021 年 9 月

東京理科大学 工学研究科 電気工学専攻

大島 浩嗣

List of papers

6.3.1 Journal paper

[1] K. Oshima, T. Onishi, S-J. Kim, J. Ma, M. Hasegawa, "Efficient wireless network selection by using multi-armed bandit algorithm for mobile terminals," *Nonlinear Theory and Its Applications*, IEICE, vol. 11, Issue 1, pp. 68–77, 2020.

[2] K. Oshima, T. Kobayashi, Y. Taenaka, K. Kuroda, M. Hasegawa, "Wireless network optimization method based on cognitive cycle using machine learning," *IEICE ComEX*, vol. 7, no. 7, pp. 278–283, 2018.

6.3.2 International conference

[1] K. Oshima, T. Kobayashi, Y. Taenaka, K. Kuroda, M. Hasegawa, "Autonomous Wireless System Optimization Method based on Cross-layer Modeling using Machine Learning," *IEEE International Conference on Ubiquitous and Future Networks*, pp. 239–244, 2019.

6.3.3 Domestic conference

[1] K. Oshima, T. Kobayashi, Y. Taenaka, K. Kuroda, M. Hasegawa, "Wireless Network Optimization Method based on Cognitive Cycle using Machine Learning," *IEICE Tech. Rep.*, vol. 117, no. 457, SR2017-132, pp. 103-107, 2018.

This dissertation is based on "Efficient wireless network selection by using multi-armed bandit algorithm for mobile terminals" [301, 302], by the same author, which appeared in the Nonlinear Theory and Its Applications, and IEICE ComEx, Copyright(C)2020 IEICE. The material in this paper was presented in part at the Nonlinear Theory and Its Applications, and IEICE ComEx [301,302] and all the figures of this paper are reused from them under the permission of the IEICE.

This dissertation is also include the paper from IEEE. Copyright(C)2019 IEEE. Reprinted, with permission, from [303].